Investigating domain-selective angiotensin converting enzyme inhibition and oxidative inactivation

Thesis submitted in fulfilment of the requirements for the degree

DOCTOR OF PHILOSOPHY IN CHEMICAL BIOLOGY

by

Lizelle Lubbe

LBBLIZ002

Department of Integrative Biomedical Sciences

Faculty of Health Sciences

University of Cape Town

June 2018

Supervisor: Professor Edward D. Sturrock; Co-supervisor: Professor B. Trevor Sewell

Declaration

I, Lizelle Lubbe, hereby declare that this thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that acknowledged below; except as stated below, neither the substance nor any part of the thesis has been submitted in the past, nor is being, nor is to be submitted for a degree at this University or any other University. I am now presenting the thesis for examination for the Degree of PhD. The Harvard referencing style was used.

I hereby grant the University free license to reproduce this thesis in whole or in part, for the purpose of research.

Student name: Lizelle Lubbe

Student number: LBBLIZ002

Signature: Signed by candidate

Date: 25 June 2018

I, Lizelle Lubbe, confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):

a) Ross G. Douglas, Rajni K. Sharma, Geoffrey Masuyer, **Lizelle Lubbe**, Ismael Zamora, K. Ravi Acharya, Kelly Chibale and Edward D. Sturrock. *Fragment-based design for the development of N-domain-selective angiotensin-1-converting enzyme inhibitors*. Clinical Science (London) (2014); 126 (4): 305-313. Doi: 10.1042/CS20130403

   Contribution to the publication: I carried out and analysed the data of the enzyme kinetics experiments and contributed to the writing of the manuscript. My contribution to the results is given in Chapter 2 section 2.6.5 of this thesis.

b) **Lizelle Lubbe**, B. Trevor Sewell and Edward D. Sturrock. *The influence of angiotensin converting enzyme mutations on the kinetics and dynamics of N-domain selective inhibition*. FEBS Journal (2016); 283 (21): 3941-3961. Doi: 10.1111/febs.13900

   Contribution to the publication: I was involved in the conceptualization of this study, carried out all the experimental work, analysed all the data and wrote the entire first draft of the manuscript. The full manuscript is presented in Chapter 2 (sections 2.1, 2.3.1-2.3.4, 2.6.2 and 2.6.12 of this thesis have not been published).

Student name: Lizelle Lubbe

Student number: LBBLIZ002

Signature: 

Date: 25 June 2018

Abstract

Angiotensin converting enzyme (ACE) is a zinc metalloprotease comprised of two homologous, catalytically active domains (90% active site identity and 60% sequence similarity). The C-domain is responsible for blood pressure regulation via angiotensin I cleavage while the N-domain inactivates an antifibrotic peptide Acetyl-Ser-Asp-Lys-Pro (AcSDKP). Since selective N-domain inhibition will result in AcSDKP accumulation, it shows promise for the treatment of fibrosis without affecting blood pressure. Low bioavailability, however, precludes the use of currently available N-selective ACE inhibitors in a clinical setting.

Inhibition of ACE by a phosphinic, peptidomimetic compound, 33RE, was characterized using a continuous assay with quenched fluorogenic substrate. The N-domain displayed nanomolar (K_i = 11.21±0.74nM) and the C-domain micromolar (K_i = 11 278±410nM) inhibition, thus 1000-fold selectivity. Residues predicted to contribute to selectivity based on the N-domain-33RE co-crystal structure were subsequently mutated to their C-domain counterparts. S_2 subsite mutation with resulting loss of a hydrogen bond drastically decreased 33RE affinity (K_i = 2794±156nM) yet did not entirely account for the selectivity. Additional substitution of all unique S_2' residues, however, completely abolished N-selectivity (K_i = 10 009±157nM). Interestingly, these residues do not directly bind 33RE. All mutants were therefore subjected to molecular dynamics (MD) simulations in the presence and absence of 33RE in addition to co-crystallization of 33RE with the N-domain mutant having all S_2 and S_2' residues mutated. Trajectory analyses highlighted the S_2' residues' importance in formation of a favourable interface between the ACE subdomains and thus a closed, ligand-bound complex. This was supported by X-ray crystallography and provides a molecular basis for the inter-subsite synergism governing 33RE's 1000-fold N-domain selectivity.

Enzyme kinetics were also used to study the concentration-dependent competitive inhibition and time-dependent irreversible oxidative inactivation of ACE catalysed by the Cu-Gly-Gly-His-lisinopril (CuGGHLis) metallodrug. Although both domains displayed nanomolar affinity for metallodrug binding (N-domain K_i = 44.94±1.84nM and C-domain K_i = 15.57±1.30nM), rapid and complete CuGGHLis-mediated inactivation occurred exclusively in the N-domain upon incubation with ascorbate and H_2O_2 redox co-reactants (k_2 = 59 710 M^{-1} min^{-1}). Michaelis-Menten characterization of the residual activity after partial N-domain inactivation

revealed a decreased rate for hydrolysis of a non-domain selective substrate. This suggests that although CuGGHLis binds with similar affinity to both domains, the metal-chelate is optimally orientated in the N- but not the C-domain to catalyze oxidation of residues involved in substrate hydrolysis.

The C-domain, in contrast, showed increased susceptibility to oxidative inactivation by diffuse radicals. This is of physiological significance as C-domain inactivation in normotensive individuals could result in accumulation of pro-inflammatory peptides. Since the N-domain is more heavily glycosylated, the potential role of unique glycans in diffuse radical shielding was studied using glycoprotein MD simulations. Unique C-domain solvent tunnels were identified that could increase diffuse radical access and, additionally, the mechanism whereby glycosylation contributes to ACE thermal stability was described for each site. This has implications for future ACE crystallography studies and the design of ACE-modulating agents with potential anti-inflammatory activity.

This study demonstrated the utility of combining *in vitro* and *in silico* approaches to reveal how subtle amino acid or glycosylation site differences between the highly homologous domains control dynamic behaviour. It furthermore elucidated how two inhibitors with different mechanisms of action selectively target the N-domain active site by exploiting these differences and provided valuable insight for future anti-fibrotic ACE inhibitor design.

Acknowledgements

I would like to express my deepest thanks to my supervisor Ed, who has provided me with so many opportunities throughout the years. Your guidance, critique, encouragement, humour and passion for ACE has been invaluable and made this an unforgettable journey. I am forever indebted to my co-supervisor Trevor for introducing me to the amazing world of structural biology and for helpful discussions around the dynamics of ACE. I thank Clare and Natasha for assistance with the glycan molecular dynamics, the CSIR CHPC for providing computational resources and UCT (Washkansky, Manuel and Luby Merit Scholarship, Duncan Baxter Merit Scholarship and UCT Doctoral Research Scholarship) and the National Research Foundation of South Africa for financial assistance during this degree. This work is based on the research supported in part by the National Research Foundation of South Africa (grant numbers 83202, 95250 and 111545).

I am grateful to Prof A.K. Carmona, Prof V. Dive, Dr R.K. Sharma, Prof J.A. Cowan, Prof K.R. Acharya and Dr N.F. Brás for kindly providing the fluorogenic peptides, RXP407 inhibitor, 33RE inhibitor, GGHLis metallodrug, 33RE crystal structure and zinc forcefield files, respectively.

To the present and past members of the ACE lab: Dale, Karabelo, Ross, Albert, Nailah, Elaine, Afolake, Bertus, Stephen, Emma and especially Vinasha, Palesa, Kate and Lauren, I thank you for all our escapades around the Cape and for making this such a fun, supporting work environment. And to Sylva who was always willing to help in the lab and encouraged me through all the ups and downs of PhD with a smile.

To my family: my parents Gert and Ilse, brother Werner, sister Nadia, aunt Lize, grandmother Ouma Gertruida, grandfather Oupa Paul and late grandmother Ouma Sus who always showed interest in my work and patiently supported me from a distance on this long road. I can't wait to see you again!

This would not have been possible without the grace of Jesus Christ, the Creator of ACE.

Dedicated to my loving parents

Gert and Ilse

Table of contents

List of tables

List of figures

List of abbreviations

°C	Degrees Celsius
µl	Microliter
µM	Micromolar
0fA	Fucose
0SA	Sialic acid
1,8-ANS	8-Anilino-1-naphthalenesulfonic acid
2MA	Mannose
3LB	Galactose
4YB	N-acetylglucosamine
Å	Angstrom
Abs	Absorbance
Abz	o-Aminobenzoic acid
(Abz)-FRK-(Dnp)-P-OH	(Abz)-Phe-Arg-Lys-(Dnp)-Pro-OH
Abz-Gly	o-Aminobenzoic acid-glycine
Abz-LFK(Dnp)-OH	Abz-Leu-Phe-Lys(Dnp)-OH
Abz-SDK(Dnp)P-OH	Abz-Ser-Asp-Lys(Dnp)Pro-OH
(Abz)-YRK-(Dnp)-P-OH	(Abz)-Tyr-Arg-Lys-(Dnp)-Pro-OH
ACE	Angiotensin converting enzyme
ACE2	Angiotensin converting enzyme 2
AcSDKP	N-acetyl-Ser-Asp-Lys-Pro
AD	Alzheimer's disease
AMU	Atomic mass units
Ang1-7	Angiotensin 1-7
AngI	Angiotensin I
AngII	Angiotensin II
AT_1R	Angiotensin receptor type I
AT_2R	Angiotensin receptor type II

ATCUN	Amino terminal copper and nickel binding
atm	Atmosphere
Aβ	Amyloid beta-peptide
B1R	Bradykinin receptor B1
B2R	Bradykinin receptor B2
BK	Bradykinin
bp	Basepair
CB-TE2A	1,4,8,11-Tetraazabicyclo[6.6.2]hexadecane-4,11-diacetic acid
Cdom	C-domain
CHO	Chinese hamster ovary
CPM	Carboxypeptidase M
CuGGHLis	Cu-Gly-Gly-His-lisinopril
cyclam	1,4,8,11-Tetraazacyclotetradecane
cyclen	1,4,7,10-Tetraazacyclododecane
dH_2O	Distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dnp	2,4-Dinitrophenyl
DOTA	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid
DTPA	Diethylenetriaminepentaacetic acid
DTT	Dithiothreitol
ECE	Endothelin converting enzyme
EDTA	Ethylenediaminetetraacetic acid
Endo-H	Endoglycosidase-H
FCS	Foetal calf serum
FI	Fluorescence intensity
fMLF	Formyl-Met-Leu-Phe

FPR	Formyl peptide receptor
Fuc	Fucose
G418	Geneticin
Gal	Galactose
GGH	Gly-Gly-His
GGHLis	Gly-Gly-His-lisinopril
GlcNAc	N-acetylglucosamine
GPU	Graphics processing unit
HBS	HEPES Buffered Saline
HCV	Hepatitis C virus
HEMGH	Zinc binding motif (His-Glu-Met-Gly-His)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HHL	Hippuryl-L-histidyl-L-leucine
h-IAPP	Human islet amyloid polypeptide
HIV	Human immunodeficiency virus
HL	His-Leu
HPLC	High performance liquid chromatography
HREST-BP	Hamiltonian Replica Exchange with solute tempering and biasing potential
IC_{20}	The amount of inhibitor required for 80% residual activity
IC_{50}	The amount of inhibitor required for 50% residual activity
IFN-γ	Interferon gamma
IgG	Immunoglobulin G
IL-1	Interleukin 1
IL-6	Interleukin 6
K	Kelvin
k_2	Second order rate constant
kBp	Kilobase pairs
k_{cat}	Enzymatic turnover rate of enzyme

kDa	Kilodalton
KGHK	Lys-Gly-His-Lys
K_i	Inhibition constant
KKS	Kallikrein-kinin system
K_m	Michaelis constant
LB	Luria Bertani
LH-RH	Luteinizing hormone-releasing hormone
Lis	Lisinopril
LRP-1	Low density lipoprotein receptor-related protein 1
Man	Mannose
MAPK	Mitogen-activated protein kinase
MD	Molecular dynamics
ml	Millilitre
mM	Millimolar
MMP	Matrix metalloproteinase
NBDNJ	N-butyldeoxynojirimycin
Ndom	N-domain
NeuAc	N-acetylneuraminic acid (predominant mammalian sialic acid)
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NK-1	Neurokinin-1 receptor
NLN	*N*-linked asparagine
nM	Nanomolar
NPT	Constant number of particles, pressure and temperature
ns	Nanosecond
n.s.	Not significant
NTA	Nitrilotriacetic acid
NVT	Constant number of particles, volume and temperature
PAF	Platelet-activating factor

PAGE	Polyacrylamide gel electrophoresis
PAI-1	Plasminogen activator inhibitor type 1
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PD	Parkinson's disease
PDB	Protein Data Bank (http://www.rcsb.org/pdb)
PG	Prostaglandin
PKC	Protein kinase C
PLC	Phospholipase C
pM	Picomolar
PMA	Phorbol 12-myristate 13-acetate
PMEMD	Particle Mesh Ewald molecular dynamics
PMSF	Phenylmethylsulphonyl fluoride
PNGs	Potential *N*-linked glycosylation site
ps	Picosecond
PTK	Protein tyrosine kinase
RAAS	Renin-angiotensin aldosterone system
RFU	Relative fluorescence units
R_i	Intrinsic reactivity to oxidation
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RT-PCR	Reverse transcription-polymerase chain reaction
sACE	Somatic ACE
SD	Standard deviation
SDS	Sodium dodecylsulphate

SEM	Standard error of the mean
SNFG	Symbol nomenclature for glycans
SP	Substance P
SrtA	Sortase A
T2DM	Type 2 diabetes mellitus
tACE	Testis ACE
TACN	1,4,7-Triazacyclononane
TBE	Tris-borate-EDTA buffer
TEMED	Tetramethylethylenediamine
TGF-β_1	Transforming growth factor β_1
T_m	Melting temperature
Tn5	*Trichoplusia ni*
TNFα	Tumour necrosis factor α
t-PA	Tissue plasminogen activator
VMB	Mannose
Z-FHL	Benzyloxycarbonyl-Phe-L-His-Leu

Chapter 1

Literature review

1.1 Physiological regulation of fluid homeostasis

Hypertension is a major risk factor for several disease states including stroke, cardiovascular disease and kidney disease and is considered one of the most important preventable causes of death (Danaei *et al.*, 2009). In 2010, 31.1% of the global adult population (1.39 billion individuals) were suffering from hypertension (Mills *et al.*, 2016). The burden of hypertension is especially high in middle- to low income countries and increased dramatically in the period between 2000 and 2010 from 599 million to 1.0 billion hypertensive individuals (Mills *et al.*, 2016). There is thus a great need for effective antihypertensive therapy.

Antihypertensive drug development started more than sixty years ago with the first well-tolerated antihypertensive, chlorothiazide, discovered in 1958 (Bolte *et al.*, 1958). Fluid and electrolyte homeostasis is regulated by the renin-angiotensin aldosterone system (RAAS) and the kallikrein-kinin system (KKS) (Acharya *et al.*, 2003). In the RAAS, the enzyme renin is produced by the kidneys and cleaves a liver-derived protein angiotensinogen (Figure 1.1). This results in release of the decapeptide angiotensin I (AngI) which is further processed into angiotensin II (AngII) through dicarboxypeptidase action of the angiotensin converting enzyme (ACE, EC 3.4.15.1). Upon binding of AngII to the angiotensin receptor type I (AT_1R) and stimulation of the adrenal gland, vasoconstriction, aldosterone release, vasopressin release and sodium retention ensues. Although AngII binding to the angiotensin receptor type 2 (AT_2R) has an opposing effect (Siragy, 2000), this receptor is only expressed during foetal development and declines after birth (Grady *et al.*, 1991). ACE further functions to increase blood pressure via the KKS by degrading the vasodilator bradykinin (BK). This enzyme is thus essential for blood pressure regulation and is an attractive antihypertensive target.

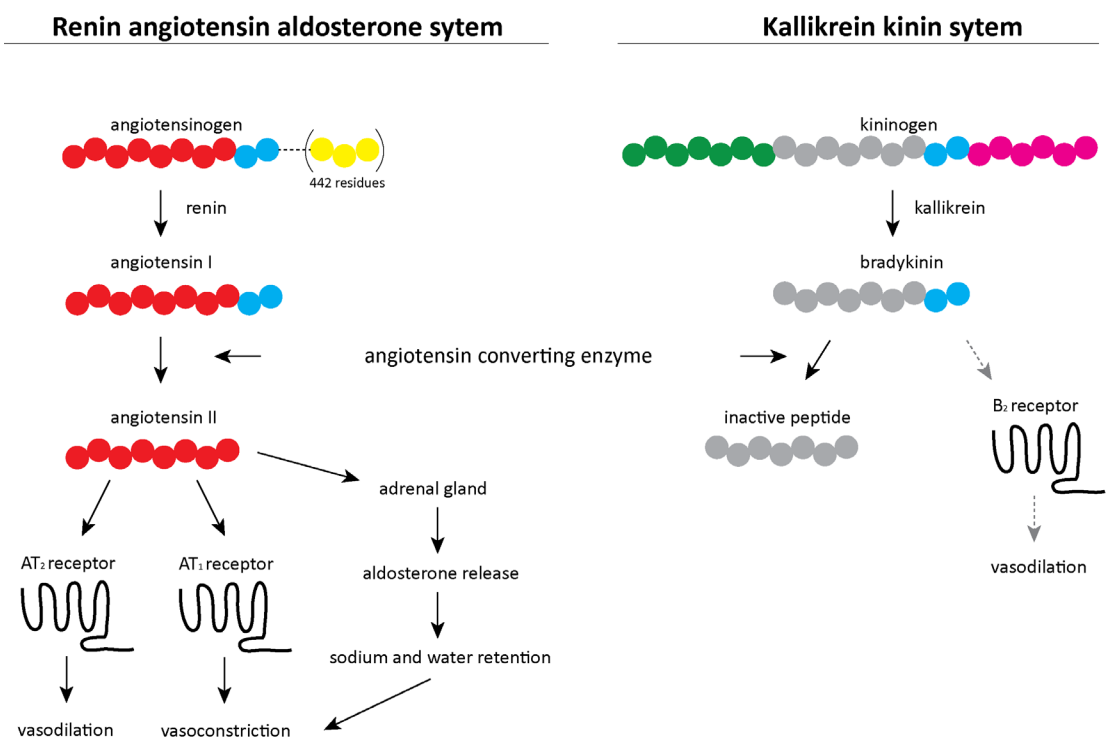


Figure 1.1: Regulation of blood pressure via the renin-angiotensin aldosterone system and kallikrein-kinin system.

1.2 Biochemistry of ACE

Transcription of a single human *Ace* gene with tissue specific promotors results in expression of two distinct isoforms (Kessler *et al.*, 2000). While the testis ACE isoform (tACE) occurs exclusively in male germinal cells and is important for fertility (Hagaman *et al.*, 1998), the somatic form (sACE) is expressed on the surface of endothelial, epithelial, neuroepithelial and immune cells and is thus widely distributed among tissues (Erdos, 1990). Release of sACE from the membrane by a yet unidentified secretase yields a soluble form which has been detected in blood, urine, cerebrospinal fluid and seminal fluid (Ehlers *et al.*, 2012). A single polypeptide chain gives rise to this type I transmembrane protein, possessing a 20-amino acid transmembrane region and 30-amino acid C-terminal cytoplasmic tail.

sACE is a zinc metalloprotease of the M2 gluzincin family and has two extracellular domains, the N- and C-domain, separated by a short linker region. Each domain contains an HEMGH zinc-binding motif and is catalytically active (Wei, 1991). In contrast, the tACE isoform consists of a single domain essentially identical to the C-domain of sACE, except for the 36-residue serine- and threonine-rich N-terminal region present in tACE (Ehlers *et al.*, 1989).

The high degree of homology between the sACE domains (60% overall sequence similarity and 89% active site identity) (Soubrier, 1988) suggests that they originated by an evolutionary gene duplication event and were conserved due to differences in their physiological function.

sACE cleaves a remarkable range of substrates through both endo- and exopeptidase action. These include AngI, enkephalins, kinins, neurotensin, formyl-Met-Leu-Phe, substance P (Erdos, 1990), gonadotropin-releasing hormone, also known as luteinizing hormone-releasing hormone (LH-RH) (Skidgel and Erdos, 1985), N-acetyl-Ser-Asp-Lys-Pro (AcSDKP) (Rousseau *et al.*, 1995) as well as the amyloid beta-peptide (Aβ) (Oba *et al.*, 2005; Zou *et al.*, 2009). The C-domain is primarily responsible for blood pressure regulation via AngI cleavage (Wei, 1991; Fuchs *et al.*, 2008) whereas the peptides LH-RH, AcSDKP and Aβ are preferentially cleaved by the N-domain (Fuchs, 2004; Rousseau *et al.*, 1995; Oba *et al.*, 2005; Zou *et al.*, 2009; Jaspard *et al.*, 1993). Although both domains hydrolyse BK, the C-domain is approximately 4-fold more efficient (Jaspard *et al.*, 1993).

1.3 ACE glycosylation

1.3.1 Sites of mammalian ACE glycosylation

sACE is present as a type-I membrane glycoprotein with ten N-domain and seven C-domain potential *N*-linked glycosylation sites (PNGs) comprised of N-X-S/T (where X is any amino acid except proline). Mass spectrometric analyses of single-domain truncated ACE expressed in mammalian CHO-K1 cells revealed that C-domain sites c1-3 (N72, N90 and N109) were fully occupied whereas sites c4-6 (N155, N337 and N586) were partially occupied and the C-terminal site c7 (N620) unoccupied (Yu *et al.*, 1997). The N-domain sites n1-6 (N9, N25, N45, N82, N117, N131) and n8-9 (N416, N480) were found to be glycosylated with the C-terminal site n10 (N494) unoccupied and the glycosylation status of site n7 (N289) unknown (Anthony *et al.*, 2010). The N-domain crystal structure (PDB ID: 2C6F), however, shows density suggesting GlcNAc attachment to N289, thus providing experimental evidence for glycosylation at this site. Based on mass spectrometric analysis of the truncated C-domain, a fucosylated sialylated biantennary complex type glycan structure was proposed (Figure 1.2) (Yu *et al.*, 1997). This was supported by earlier quantitative HPLC analysis on human kidney sACE and tACE expressed in CHO-K1 cells (Ehlers *et al.*, 1992).

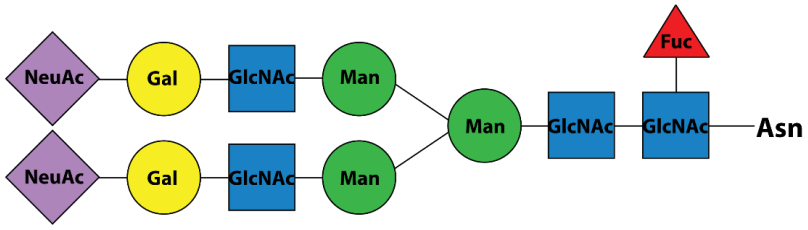


Figure 1.2: Proposed structure of the glycan attached to ACE PNGs.
Prepared using the GlycoBuilder Tool contained within the GlycoWorkbench v.2.1 suite (Ceroni *et al.*, 2008) and represented with standard Symbol Nomenclature for Glycans (SNFG) notation (Varki *et al.*, 2015).

1.3.2 Influence of glycosylation on ACE folding and function

Glycosylation is involved in glycoprotein folding, localization, stabilization, activity, receptor binding and dimerization (Mitra *et al.*, 2006). In ACE the glycan locations and functions are domain-specific, apart from N-domain sites n3, n4 and n6 having equivalent locations to C-domain sites c1, c3 and c4. The lack of non-glycosylated C-domain secretion into the medium has been reported, suggesting a role for glycosylation in protein expression and processing (Gordon *et al.*, 2003). Previously, it was shown that secretion of the amyloid precursor protein is enhanced by sialylation at its *N*-linked sites (Nakagawa *et al.*, 2006). Secretion of ACE, however, does not appear to rely on sialic acid protein interactions since glycosylation with $Man_5GlcNAc_2$ (CHO-K1 glycosylation mutant *Lec*3.2.8.1 or *Trichoplusia ni* (*Tn5*) expression) or $Glc_3Man_{7\text{-}9}GlcNAc_2$ (CHO-K1 expression in presence of N-butyldeoxynojirimycin (NBDNJ)) still yielded correct folding of N- and C-domain (Redelinghuys, 2006; Gordon *et al.*, 2003; Yu *et al.*, 1997; O'Neill *et al.*, 2008).

The contribution of glycosylation at specific sites towards protein folding was previously investigated by N>Q mutation of the truncated N- and C-domain PNGs. It was concluded that correct N-domain folding and secretion occurs in the presence of glycosylation at C-terminal sites n7-9 or a combination of sites n1, n2 or n3 and n7, n8 or n9 (Anthony *et al.*, 2010) (Figure 1.3). Glycosylation at two C-terminal (Ndom7,9 or Ndom8-9) or only N-terminal (Ndom1; Ndom3; Ndom1,3; Ndom2-3; Ndom1-3; Ndom1-4; Ndom1-5; Ndom1-6) sites resulted in protein misfolding and rapid intracellular degradation (Anthony *et al.*, 2010; Redelinghuys, 2006). C-domain folding, on the other hand, could be achieved with glycosylation at sites c1; c1-2; c3; c1 and c3; c1-3; c2-3; or c1-4. Site c1 alone slightly decreased the efficiency of

processing and solubilization whereas site c2 alone resulted in misfolding and rapid intracellular degradation (Gordon *et al.*, 2003).

Although glycosylation is important in folding and processing of the truncated domains, it does not appear to play a role in enzymatic function of mature protein. N-domain catalytic efficiency between CHO-K1, CHO-K1 *Lec3.2.8.1* or CHO-K1+NBDNJ expressed protein (Redelinghuys, 2006) and specific activity between glycosylation mutants and wild-type is comparable (Anthony *et al.*, 2010). Similarly, C-domain catalytic efficiency is not affected by treatment with NBDNJ, glycosylation site mutation (Gordon *et al.*, 2003) nor removal of all but a single GlcNAc residue at each site (after endoglycosidase-H (Endo-H) treatment of CHO-K1+NBDNJ expressed protein) (Yu *et al.*, 1997).

1.3.3 The role of glycosylation in thermal stability

Apart from mediating folding, glycans can also function in stabilizing the underlying protein by directly interacting with certain surface residues (Mitra *et al.*, 2006; Lee *et al.*, 2015). This could increase the protein's resistance to thermal degradation. The N-domain melting temperature (T_m) is markedly higher than that of the C-domain and even allows purification of the single N-domain by selective thermal degradation of the C-domain (O'Neill *et al.*, 2008; Voronov *et al.*, 2002a). Earlier reports suggested that the N-domain's superior thermal stability is due to its N-terminal region's high α-helicity, proline content and glycosylation (Voronov *et al.*, 2003). Site-directed mutagenesis of the N-X-S/T sites to Q-X-S/T, however, revealed that glycosylation at C-terminal regions is also involved.

Thermal stability is conferred by complex-type sugars at N-domain sites n2 and n3 together with n7 or n9 with markedly increased thermal degradation upon loss of either site n2 or n3 (O'Neill *et al.*, 2008; Anthony *et al.*, 2010) (Figure 1.3). This stabilization might involve sialic acid interactions with the protein since the N-domain (in sACE form) T_m decreased by 6°C when expressed in *Tn5* insect cells (lacking sialic acid residues) instead of CHO-K1 cells (O'Neill *et al.*, 2008). As opposed to loss of direct interactions it could, however, also be related to an increased solvent accessible surface area upon glycosylation with the smaller $Man_5GlcNAc_2$. Since not only glycan-protein, but also glycan-glycan interactions occur, glycosylation could significantly decrease the exposed protein surface area. Previously, Kryukova *et al* reported a contact area of 200–300Å between ACE and a sialylated

biantennary complex glycan (Kryukova *et al.*, 2015). Indeed, the C-domain T_m was decreased by only 2°C with CHO-K1+NBDNJ expression ($Glc_3Man_{7-9}GlcNAc_2$) but by 6°C when expressed in *Tn5* cells ($Man_5GlcNAc_2$), as seen for the N-domain. Since removal of sites c5-6 or c2, c4, c5 and c6 decreased C-domain T_m by 6°C and 8°C, respectively, stabilization is largely conferred by sites c5 and c6 (O'Neill *et al.*, 2008). At present, however, the exact mechanism whereby glycosylation protects ACE against thermal denaturation and the basis for N-domain superiority is unknown.

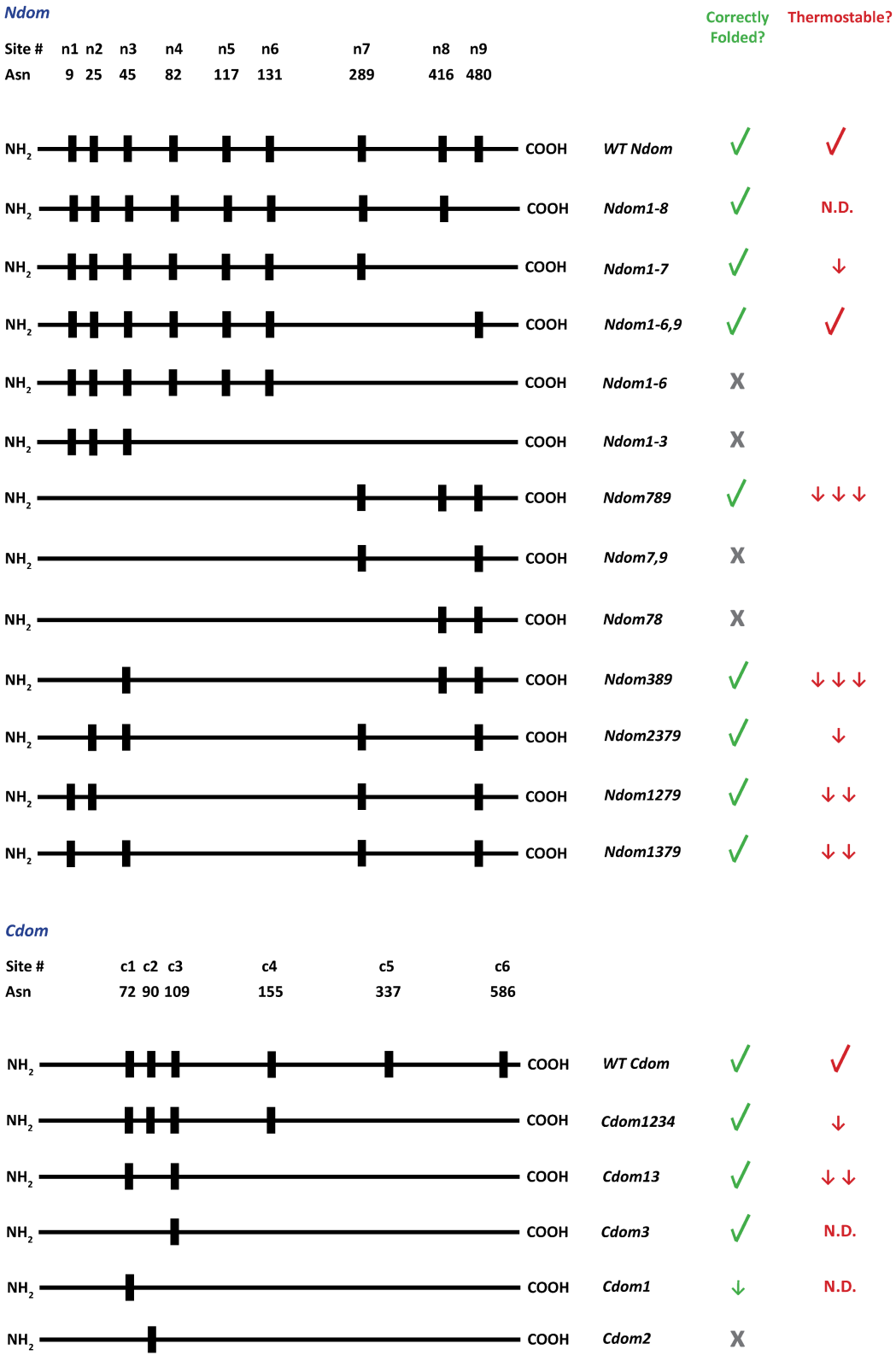


Figure 1.3: Glycan contribution to ACE folding and thermal stability.
C-domain numbering is given as standard tACE numbering (N.D.: T_m not determined; arrows signify the degree by which T_m was decreased; crosses indicate aberrant protein folding and resulting intracellular degradation).

1.4 ACE structural biology

1.4.1 Structures of the N- and C-domain of ACE

The degree of sACE glycosylation (30% carbohydrates by weight) presented a major hurdle for structural studies using X-ray crystallography since the flexibility of glycans is detrimental for crystal packing. While enzymatic deglycosylation might yield purified protein amenable to crystallization, such an approach is costly and potential sample loss could be problematic for a technique which requires large protein quantities. The first ACE structure was therefore solved using a site-directed mutagenesis approach to obtain a truncated C-domain protein with the minimum glycan requirement for correct folding and processing (Natesh *et al.*, 2003). Subsequently, the truncated N-domain structure was also solved using this approach (Anthony *et al.*, 2010). The C-domain was glycosylated at sites c1 and c3 while the N-domain contained glycans at sites n3, n8 and n9.

The overall topology is very similar between the two domains, in line with their high sequence similarity (Corradi *et al.*, 2006). Each ellipsoid ACE domain is largely composed of alpha helices which are separated by several flexible loops and six beta sheets (Figure 1.4). The tertiary structure is further formed by three cysteine disulphide bridges. The single unbound cysteine is located on a parallel beta sheet on the surface whereas two of the three cysteine disulphide bridges are proximal to the anti-parallel beta sheets. The active site zinc is located within a large, occluded cleft which divides each domain into two subdomains and is covered by a lid-like structure comprised of the first three N-terminal helices. Based on the lid's high B-factors, it has been suggested that hinging in this region could mediate opening of the active site cleft to provide access to the large range of ACE substrates (Watermeyer *et al.*, 2006; Anthony *et al.*, 2010). Catalytic efficiency of especially the C-domain is dependent on the presence of chloride (Yates *et al.*, 2014). Some insight into the higher C- than N-domain chloride sensitivity was provided by the presence of two chloride ions in the C-domain and one chloride in the N-domain structure.

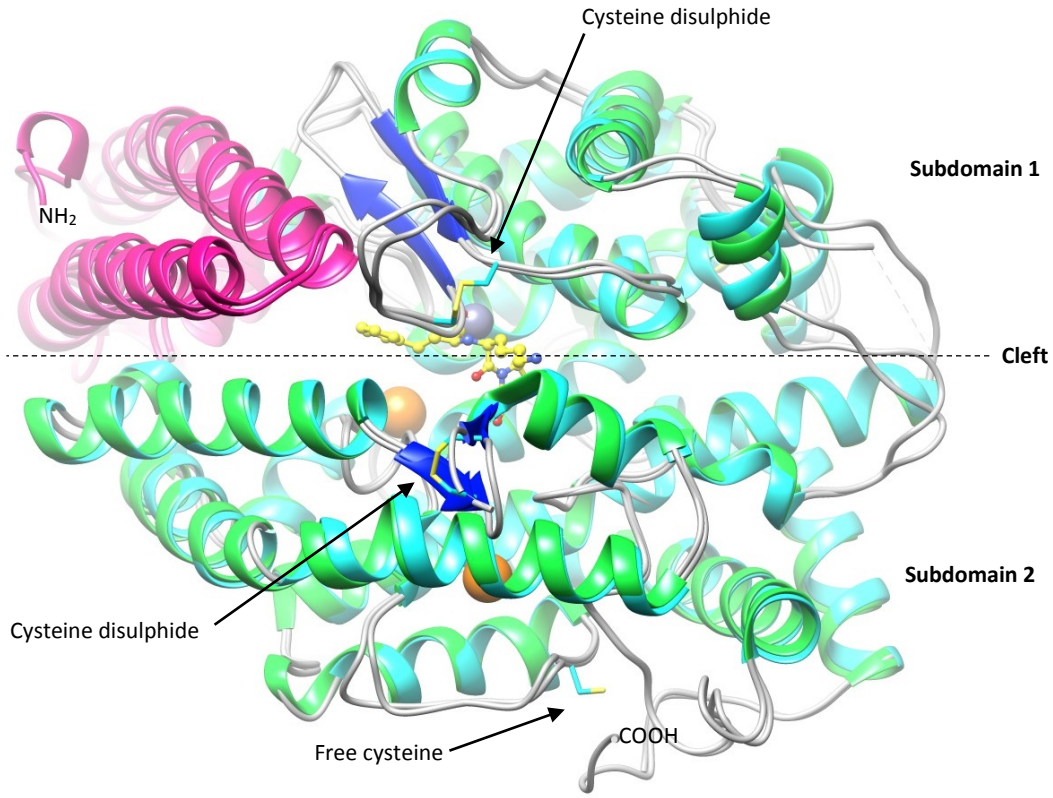


Figure 1.4: Structural alignment of the N- and C-domain of ACE.
The N-domain (PDB ID: 3NXQ) and C-domain (PDB ID: 1O86) helices are given in green and cyan, respectively, with the N-terminal lid formed by helices 1-3 coloured dark pink. The cysteine residues are shown as cyan sticks and the chloride ions as orange spheres. The active site zinc ion within the deep cleft (dotted line) is shown as a grey sphere complexed to the ACE inhibitor, lisinopril (yellow ball-and-stick representation).

1.4.2 Structure of the two-domain sACE

Physiologically, however, sACE is comprised of both domains and their interactions have been shown to cooperatively affect not only its catalytic efficiency, but also inhibitor binding (Skirgello *et al.*, 2005). Furthermore, it has been proposed that upon inhibitor binding, intracellular signalling is activated via sACE dimerization to ultimately stimulate ACE expression (Kohlstedt *et al.*, 2004; Kohlstedt *et al.*, 2006). To gain insight into the two domain's relative orientations and the mechanism whereby these processes occur, the full-length sACE structure is required. Use of the mentioned PNGs N>Q mutagenesis approach for sACE, however, failed to produce correctly folded, minimally glycosylated full-length protein amenable to crystallization (Nkoe, 2014; Anthony, 2011). Removal of glycosylation sites led to intracellular cleavage at the inter-domain linker and subsequent degradation of the individual domains. Although three-dimensional electron microscopic reconstruction of

negatively stained porcine sACE has been performed (Chen *et al.*, 2010), fitting of the truncated domains into this low resolution (23Å) structure did not offer much atomic detail. At present, the precise orientation of the two domains in human sACE form and the location of inter-domain contacts is therefore not known (Corradi *et al.*, 2006). As such, the mechanisms of inter-domain substrate cooperativity and sACE homodimerization remain elusive. Understanding the functional role of each domain's glycosylation sites could greatly enhance the likelihood of successfully obtaining a correctly folded, minimally glycosylated sACE variant amenable to crystallization.

1.5 First generation ACE Inhibitors

Isolation of bradykinin potentiating peptides (BPPs) from *Bothrops jararaca* snake venom led to design of the first orally effective ACE inhibitor, captopril (Ondetti *et al.*, 1977). The first generation of clinically approved compounds were developed based on the structure of thermolysin and carboxypeptidase A, thought to be related to ACE (Ondetti *et al.*, 1977; Patchett *et al.*, 1980). Captopril displayed low nanomolar potency and effectively lowered blood pressure by preventing the production of AngII and breakdown of BK. It was therefore approved for clinical use in 1981 but later associated with the adverse side-effects of skin rash and loss of taste (Acharya *et al.*, 2003). This was attributed to the presence of a sulfhydryl zinc-binding group. To circumvent this, subsequent ACE inhibitors were designed to rather contain ketone (keto-ACE and ketomethylene analogues), carboxylate (lisinopril, enalaprilat and ramiprilat) or phosphonate (fosinoprilat) metal-chelating moieties (Redelinghuys *et al.*, 2005) (Figure 1.5).

A further benefit of these inhibitors over captopril was that they interacted with both the prime and non-prime subsites flanking the scissile peptide bond (Acharya *et al.*, 2003). This enabled the design of drugs with increased specificity over other metalloproteases by incorporating weaker zinc-binding groups and optimizing interactions to the ACE subsites. The current clinically available ACE inhibitors are, however, unable to distinguish between the two domains of ACE and result in systemic bradykinin accumulation. In a subset of patients this has been associated with persistent cough (5-20% of patients) and potentially life-threatening angioedema (0.1-0.2% of patients) (Bicket, 2002).

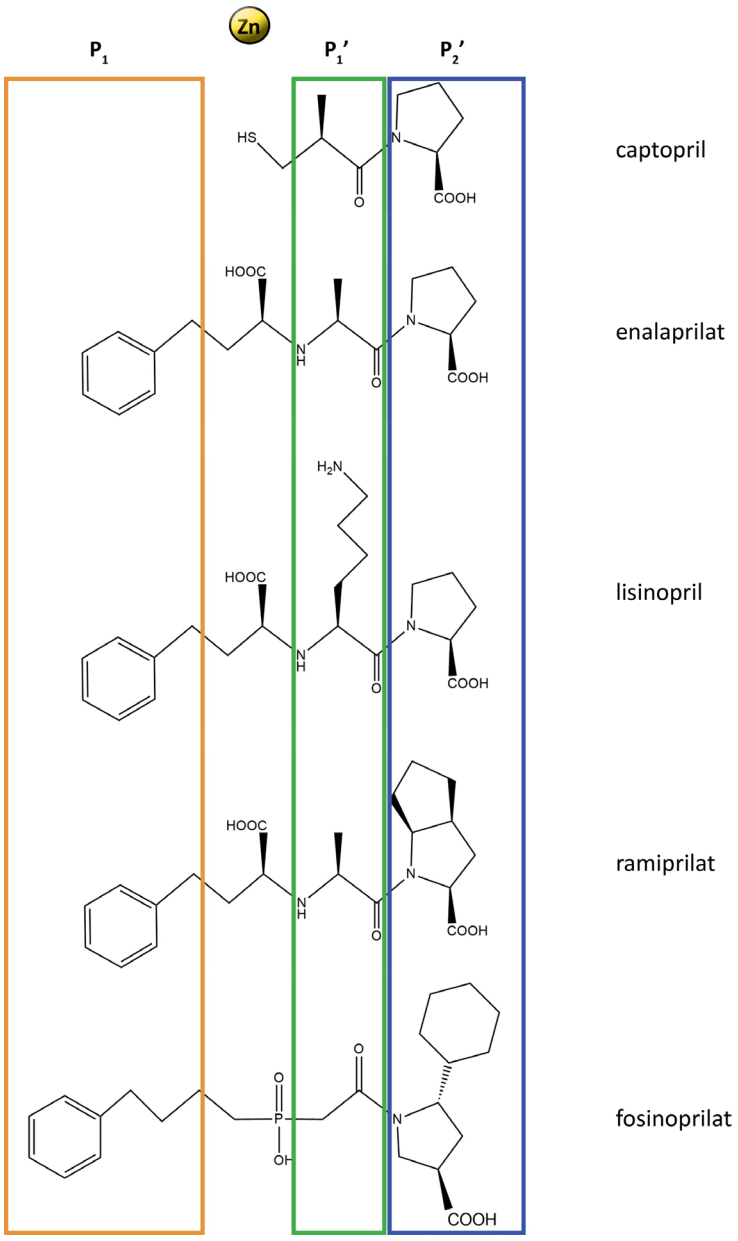


Figure 1.5: Chemical structures of ACE inhibitors commonly used in the treatment of hypertension. The zinc binding groups are flanked by the P_1, P_1' and P_2' moieties indicated by the orange, green and blue boxes, respectively.

1.6 Second generation C-domain selective inhibitors

Following reports of these adverse side-effects it was therefore proposed that the next generation of ACE inhibitors should be designed to selectively inhibit the C-domain, leaving the N-domain free to hydrolyse BK (Acharya *et al.*, 2003). Keto-ACE and the 11-mer naturally occurring bradykinin potentiating peptide b (BPPb) respectively displayed approximately 40- and 260-fold selectivity for binding to the C-domain (Deddish *et al.*, 1998; Cotton *et al.*, 2002). Based on the observation that constrained, ACE inhibitors such as BPPs displayed more C-selectivity, Georgiadis *et al* synthesized a series of pseudo proline-containing phosphinic peptide inhibitors (Georgiadis *et al.*, 2003; Georgiadis *et al.*, 2004). The RXPA380 compound displayed low nanomolar affinity and an astounding 3300-fold selectivity for binding to the C-domain. Shortly thereafter, a turning point came in the design of ACE inhibitors with the structural determination of tACE in complex with lisinopril (Natesh *et al.*, 2003). Important insights were gained into the factors responsible for potency of inhibition and, more importantly, unique residues that could be targeted in C-selective inhibitor design were identified. This led to improvements in keto-ACE selectivity (Nchinda *et al.*, 2006b; Sharma *et al.*, 2012) and development of a C-selective tryptophan analogue of clinically used lisinopril, Lisinopril-Trp (Nchinda *et al.*, 2006a).

1.7 ACE inhibition and fibrosis

Apart from decreasing blood pressure, the clinical use of non-selective ACE inhibitors has been associated with various antifibrotic effects. The fibrinolytic system functions to promote wound healing via an inflammatory response which triggers recruitment and differentiation of fibroblasts with subsequent synthesis of extracellular matrix proteins (Wynn, 2008). Conditions such as cardiovascular disease, insulin resistance, diabetes and obesity, however, can lead to prolonged inflammation and disruption of the normal tissue repair system (Ghosh and Vaughan, 2012). This results in the hallmarks of fibrosis, namely increased synthesis of extracellular matrix proteins, progressive tissue scarring, decreased tissue elasticity and impaired organ function (Wynn, 2008). An estimated 45% of human deaths are due to fibroproliferative disorders such as fibrosis (Bitterman and Henke, 1991) with lung fibrosis having a mean survival period of 2-4 years (du Bois, 2011). A recent systematic review suggested that the incidence of idiopathic pulmonary fibrosis is on par with that of stomach, liver, testicular and cervical cancers (Hutchinson *et al.*, 2015). At present, there is no effective treatment for fibrosis.

1.8 Antifibrotic mechanism of ACE inhibition

AngII is known to increase fibroblast proliferation, inflammation and transforming growth factor-β_1 (TGF-β_1) signalling (Wynn, 2008). Inhibition of ACE could thus decrease fibrosis by limiting AngII production. Infusion with the N-domain selective antifibrotic peptide AcSDKP, however, resulted in anti-inflammatory and antifibrotic effects equal to that of non-domain selective ACE inhibition (Rasoul *et al.*, 2004). Since co-administration of a monoclonal antibody against AcSDKP significantly reduced these effects of ACE inhibition, they are primarily mediated by AcSDKP (Peng *et al.*, 2005). Administration of this peptide either before or after fibrosis was shown to attenuate TGF-β_1 signalling and AT_1 receptor expression, thereby decreasing collagen deposition and myofibroblast differentiation (Xu *et al.*, 2012). Since AcSDKP is ubiquitously expressed and selectively hydrolysed by the N-domain of ACE (Rousseau *et al.*, 1995; Fuchs, 2004), selective N-domain inhibition holds promise for the treatment of fibrosis without affecting blood pressure.

1.9 Second generation N-domain selective inhibitors

Prior to solving the ACE structure, Dive *et al* screened a phosphinic peptide library and discovered a compound, named RXP407, which displayed 2083-fold selectivity for binding to the N-domain (Dive *et al.*, 1999). It is, however, unsuitable for therapeutic use due to its large size, low bioavailability and peptidic nature. Peptidic inhibitors are generally considered poor drug candidates because of their low oral bioavailability, poor solubility and susceptibility to degradation by proteases (Redelinghuys *et al.*, 2005). Functional groups essential for N-selectivity therefore need to be identified to enable the design of clinically suitable N-domain inhibitors.

RXP407 has N-terminal acetylated aspartate (P_2), L-pseudophenylalanine (P_1), L-pseudoalanine (P_1') and C-terminal amidated alanine (P_2') moieties that respectively bind to the S_2, S_1, S_1' and S_2' subsites of the ACE active site (Figure 1.6). Previous studies have shown a marked decrease in binding selectivity when the P_2 and P_2' moieties were altered, suggesting key interactions with residues from these active site pockets (Dive *et al.*, 1999). Mutation of the N-domain S_2 residues Y369 and R381 to their corresponding C-domain counterparts (phenylalanine and glutamic acid, respectively) also led to markedly decreased inhibitor binding affinities (Kroger *et al.*, 2009). Considering this, a fragment based design

approach was recently employed to substitute the P_2 acetylated aspartate of RXP407 by chemical entities predicted to interact similarly with these unique N-domain S_2 residues (Douglas *et al.*, 2014). Molecular docking to the N-domain crystal structure suggested that one analogue, 33RE, could mimic RXP407 (Figure 1.6).

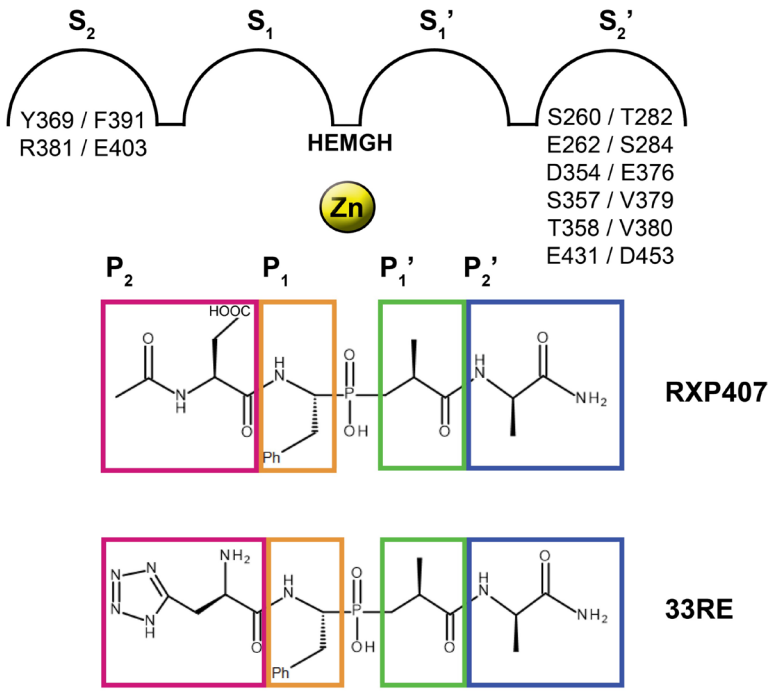


Figure 1.6: Schechter-Berger diagram illustrating the S_2 to S_2' enzyme subsites flanking the catalytic zinc-binding motif (HEMGH) binding to the P_2 to P_2' moieties of N-domain inhibitors RXP407 and its analogue 33RE.
The unique S_2 and S_2' subsite residues are indicated in N-/C-domain nomenclature.

Following synthesis and co-crystallization of 33RE with the N-domain, it was observed that the S_1 to S_2' interactions were conserved (Douglas *et al.*, 2014). Replacement of the P_2 acetylated aspartate with an aminomethyl tetrazole moiety, however, led to loss of interaction with the unique R381 so that the only unique N-domain interaction of 33RE was to Y369. Although replacement of the acid yielded a more drug-like molecule and simplified synthesis, kinetic characterization of 33RE is thus required to determine if this occurred at the expense of N-selectivity. The mere presence of a P_2 acetylated aspartate, however, does not guarantee N-selectivity as recently evidenced by the P_2 aspartate analogue of sampatrilat's lack of selectivity (Cozier *et al.*, 2018). The P_1 to P_2' moieties of RXP407 and 33RE therefore also play a role but, since their interactions are conserved between the

domains, the mechanism is poorly understood. As such, no N-selective ACE inhibitors are clinically available at present.

1.10 Irreversible inhibition of ACE

In addition to the adverse side effects of persistent cough and angioedema, clinically available ACE inhibitors such as lisinopril, benazepril, moexipril and fosinopril also suffer from poor oral bioavailability (Brown and Vaughan, 1998). The daily dosage required for effective hypertension treatment is generally high with that of captopril, lisinopril, benazepril, moexipril and fosinopril at 6.25-300 mg, 5-40 mg, 5-80 mg, 7.5-30 mg and 10-80 mg, respectively. These inhibitors compete with substrates for binding to the active site of ACE and as such only offer transient, reversible inhibition. Pseudo mechanism-based inactivators, on the other hand, have the potential to permanently deactivate the enzyme via covalent bond formation. In an attempt to obtain ACE-inhibitory compounds with a longer duration of action, Choo *et al* designed two series of epoxyalkanoyl and α, β-unsaturated carbonyl analogues (Choo *et al.*, 1998; Park Choo *et al.*, 2000). Some compounds irreversibly inactivated ACE, presumably by forming a Michael adduct with the zinc binding motif's glutamate. Their binding affinity, however, was significantly weaker than that of competitive ACE inhibitors with the most potent inactivator having an IC_{50} of 0.23mM.

1.11 Irreversible target inactivation

An alternative means of irreversible enzyme inactivation is through catalytic metallodrugs. These compounds characteristically have a target-recognition domain coupled via a flexible linker to a metal-binding domain (Cowan, 2008). The target-recognition domain is usually a ligand which binds with moderate affinity to the target of interest. Artificial nucleases are catalytic metallodrugs targeting DNA or RNA while artificial proteases target peptides or proteins. The metal-binding domain consists of a chelator complexed to a transition metal which facilitates the conversion of dioxygen or hydrogen peroxide oxidants into reactive oxygen species (ROS) through Fenton-type chemistry (Figure 1.7). The produced ROS subsequently results in cleavage of the target nucleic acid, peptide or protein and thus irreversible inactivation. Coupling of these two domains ensures localized, as opposed to diffuse, ROS production and target cleavage at specific sites (Cowan, 2008). Additionally, multiple turnover reactions occur since oxidized metal chelates are regenerated in the

presence of cellular reductants such as ascorbic acid or NADH (Gokhale and Cowan, 2005; Gokhale and Cowan, 2006; Joyner *et al.*, 2012; Fidai *et al.*, 2014). Inherently, the likelihood of toxicity and side-effects is lower for catalytic metallodrugs since their regeneration upon target inactivation allows efficacy at sub-stoichiometric concentrations and thus lower or less frequent dosing. This offers an advantage over irreversible suicide inhibitors which are consumed upon covalent modification of the target and thus required at saturating concentrations for efficacy. Since the advent of drug resistance, there has been a great need to obtain pharmaceuticals with novel mechanisms of action. Incorporation of transition metals not only offers a novel catalytic mechanism but also greatly enhances the chemical diversity achievable in drug design (Agbale *et al.*, 2016).

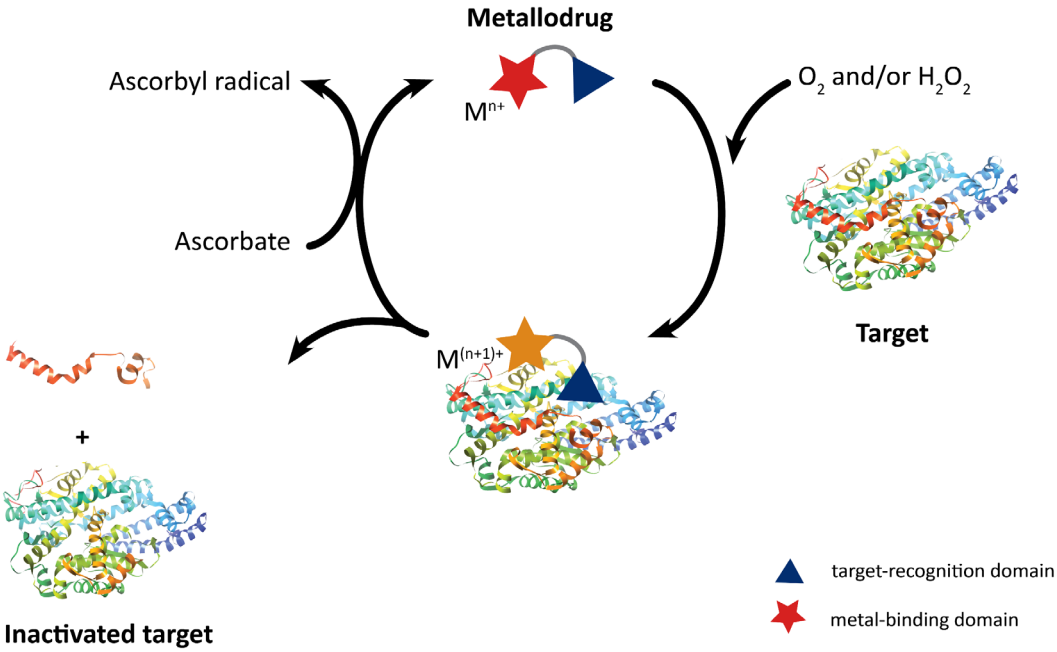


Figure 1.7: Irreversible target inactivation by metallodrug-catalysed Fenton-type chemistry. DNA, RNA, peptide or protein targets can be irreversibly inactivated through oxidative cleavage. Metallodrug dissociation from the inactivated target and regeneration by reduction in the presence of ascorbate enables efficacy at sub-stoichiometric drug concentrations. M^{n+}: transition metal complexed to the metallodrug's metal-binding domain.

1.11.1 Artificial nucleases

In 1966, a natural molecule named bleomycin was isolated from *Streptomyces verticillus* and found to be comprised of DNA-binding, metal-binding and carbohydrate domains (Umezawa

et al., 1966). It was initially recognized for its antibiotic activity but subsequently approved for the treatment of Hodgkin's lymphoma, squamous cell carcinoma and testicular, cervical and ovarian cancers (Yu and Cowan, 2017). Metal-bound bleomycin is believed to promote cancer cell apoptosis by intercalating to DNA and inducing strand breaks via ROS generation (Galm *et al.*, 2005). Currently, bleomycin is the only catalytic metallodrug which has been approved for clinical treatment of a disease. A major hurdle for the clinical application of catalytic metallodrugs has been their lack of substrate specificity *in vivo* (Yu and Cowan, 2017). Though numerous compounds developed in recent decades displayed high catalytic efficiency *in vitro*, their inability to differentiate between different cellular targets would result in decreased catalytic efficiency *in vivo*. Furthermore, off-target catalytic inactivation can result in unwanted side-effects or toxicity. The Fe^{2+}-bleomycin complex, for example, is moderately selective for binding and cleavage around GC sequences in the DNA (Sugiyama *et al.*, 1986). Due to the common occurrence of these sequences, however, cancerous as well as normal cells are targeted and treatment is thus associated with side-effects. This prompted inorganic chemists to rationally design catalytic metallodrugs with high chemical reactivity as well as target selectivity by optimizing the target-recognition and metal-binding domains employed. An ideal compound should display low reactivity in the absence of target recognition, but rapid ROS generation upon binding to the target and optimal orientation of the metal-binding domain (Cowan, 2008). By coupling different combinations of transition metals (Fe^{2+}, Co^{2+}, Ni^{2+} or Cu^{2+}) and chelators (cyclam, cyclen, CB-TE2A, DOTA, DTPA, EDTA, NTA, TACN, tripeptide Gly-Gly-His or tetrapeptide Lys-Gly-His-Lys) to nucleic acid-targeting molecules, compounds with therapeutic potential for human immunodeficiency virus (HIV) (Jin and Cowan, 2006; Joyner and Cowan, 2011), hepatitis C virus (HCV) (Bradford and Cowan, 2012) and cancer (Yu *et al.*, 2015) have since been developed.

1.11.2 Artificial proteases

Artificial proteases were initially developed to address the limited number of selective protein-cleaving agents available for studying protein sequence, structure and function (Hoyer *et al.*, 1990; Schepartz and Cuenoud, 1990). Early research was focused on understanding the mechanism whereby streptavidin (Hoyer *et al.*, 1990), calmodulin (Schepartz and Cuenoud, 1990; Cuenoud *et al.*, 1992) and carbonic anhydrase (Gallagher *et al.*, 1998) were cleaved since structural information and target-recognition domains were available for these proteins. Although the therapeutic potential of artificial proteases was already realized in 1990 (Hoyer *et al.*, 1990), the first metallodrug against a disease-related

protein was only discovered fourteen years later. By synthesizing a library of target-recognition domain candidates coupled to a Co^{3+}-cyclen complex, Chae *et al* obtained one compound which catalysed rapid and selective cleavage of the peptide deformylase protein (Chae *et al.*, 2005). This is an attractive target in antibiotic drug design due to its role in prokaryotic protein translation (Chae *et al.*, 2005). More recently, catalytic metallodrugs have also been designed for the treatment of *Staphylococcus aureus* infection by targeting the Sortase A (SrtA) protein to potentially interrupt bacterial and host cell interaction (Fidai *et al.*, 2014). This Gram-positive bacterium is a major cause of hospital-acquired infections and has developed resistance against various standard antibiotics. The designed Cu^{2+}-Gly-Gly-His (Cu^{2+}-GGH) metallodrugs significantly reduced SrtA activity by selective sidechain oxidation of critical active site residues. In this case, protein cleavage via protein backbone oxidation did not occur.

Catalytic metallodrugs can also be designed to target peptides or proteins lacking active sites. Type 2 diabetes mellitus (T2DM), Alzheimer's disease (AD), Parkinson's disease (PD), Transmissible spongiform encephalopathies and Huntington's disease are thought to respectively arise from the presence of soluble oligomers of human islet amyloid polypeptide (h-IAPP), amyloid-β_{42} ($A\beta_{42}$), α-synuclein, prion and polyglutamine amyloidogenic peptides or proteins (Haataja *et al.*, 2008). Co^{3+}-cyclen metallodrugs specifically targeted to these peptides or proteins and efficiently catalysing their cleavage have shown great promise for the treatment of T2DM, AD and PD (Suh *et al.*, 2007; Suh *et al.*, 2008; Lee *et al.*, 2008; Chei *et al.*, 2011). Apart from catalysing peptide bond cleavage, Co^{3+}-cyclen metallodrugs can also drive oxidative decarboxylation of N-terminal aspartate-containing peptides (Kim *et al.*, 2005). This has been demonstrated for the melanin-concentrating hormone and could have implications for the treatment of obesity (Kim *et al.*, 2006). Since AngII contains an N-terminal aspartate and is associated with hypertension, Co^{3+}-cyclen metallodrugs targeting AngI or AngII were also designed and two compounds subsequently identified to catalyze their oxidative decarboxylation (Kim *et al.*, 2005). At present, however, the implications for hypertension treatment are unclear since further studies are required to verify if the conversion of aspartate to pyruvate would prevent binding of AngI or AngII to ACE and AT_1, respectively.

Anti-hypertensive catalytic metallodrugs have also been designed to target ACE. The tetrapeptide Lys-Gly-His-Lys (KGHK) complexed to Cu^{2+} displayed competitive micromolar inhibition and irreversible catalytic inactivation of ACE at sub-stoichiometric concentrations in the presence of ascorbate (Gokhale and Cowan, 2005). Since the endothelin converting enzyme (ECE), which regulates blood pressure via the endothelin system, was also inactivated by this complex, this can potentially be applied as a dual-action therapeutic (Gokhale and Cowan, 2006). The clinical use of $[KGHK\text{-}Cu]^{+}$, however, will likely result in side-effects due to its DNA-cleavage activity (Jin and Cowan, 2005). Since ECE contributes to bradykinin hydrolysis (Hoang and Turner, 1997), simultaneous inactivation of ACE and ECE is further expected to exacerbate the side-effects associated with its accumulation. Although this approach appears promising for targeting ACE, optimization of the target-recognition domain would be required to circumvent these side-effects. Tailoring it towards the C- or N-domain could further enable the specific treatment of hypertension or fibrosis at sub-stoichiometric concentrations by irreversible inactivation. Catalytic metallodrugs possess so-called double-filter selectivity where ROS is only produced if the target-recognition domain is bound and the metal-binding domain optimally orientated towards readily oxidizable amino acids in the protein (Figure 1.8). This novel mechanism of action might be more sensitive to the slight variations between the ACE domains' active site residues (90% identity) than traditional competitive ACE inhibitors.

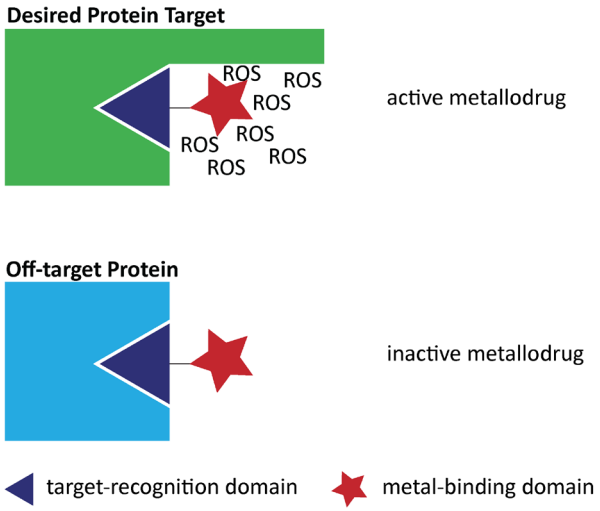


Figure 1.8: Double-filter selectivity displayed by the ideal catalytic metallodrug.
The coupled target-recognition and metal-binding domains are carefully tailored during metallodrug design to ensure selective binding to the target protein and, even if off-target binding occurs, selective localized ROS production only at the desired site.

Their decreased effective dose further lessens the importance of high bioavailability for drug approval. Altogether, this might increase the probability of obtaining a domain-selective inhibitor with pharmacokinetics suitable for clinical use. Great strides have been made in this regard by coupling a range of chelates and transition metals to the lysine sidechain of the competitive inhibitor lisinopril (Joyner *et al.*, 2012). An overview of these metallodrugs will be given in Chapter 3.

1.12 Aims and objectives

Despite availability of the truncated N- and C-domain crystal structures, factors governing selective inhibitor binding are uncertain. Additionally, efforts at obtaining the full-length sACE structure are hampered by a limited understanding of its *N*-glycans' function. The current study therefore aimed to provide a detailed description of each glycan site's functional role as well as insight into the molecular mechanisms whereby selective N-domain inhibition and catalytic inactivation occurs.

The objectives were to:

1. Identify the active site residues contributing to N-selective inhibition and elucidate their molecular mechanism using site-directed mutagenesis, enzyme kinetics, X-ray crystallography and molecular dynamics simulations, with a focus on the phosphinic ACE inhibitor 33RE

2. Study the mechanism of N-selective, metallodrug-mediated catalytic inactivation using enzyme kinetics

3. Investigate the contribution of each N- and C-domain glycan site towards protein stability by performing glycoprotein molecular dynamics simulations

Chapter 2

Elucidating the molecular basis of N-selective ACE inhibition

2.1 Introduction

The N-domain inhibition is intriguing since it prevents cleavage of AcSDKP and leads to its accumulation without affecting blood pressure (Junot *et al.*, 2001) (see Chapter 1 section 1.8). Accumulation of AcSDKP has been linked to decreased deposition of collagen and proliferation of fibroblasts in the left ventricle, kidney and lungs (Yang *et al.*, 2004; Xu *et al.*, 2012; Peng *et al.*, 2003; Liao *et al.*, 2010). As such, it could have clinical importance in the treatment of fibrosis. The adverse effects currently associated with non-selective ACE inhibitors are unlikely to occur with N-selective inhibition since the C-domain would be free to hydrolyse BK (Bicket, 2002).

To design new chemical entities for therapeutic intervention, the functional groups essential for N-selectivity should be identified. Previous studies have shown a marked decrease in RXP407's 2083-fold N-selectivity when the P_2 moiety was altered or the N-domain S_2 residues (Y369 and R381) mutated to their corresponding C-domain counterparts (Phe and Glu, respectively) (Dive *et al.*, 1999; Kroger *et al.*, 2009). Since these mutations did not completely abolish selectivity, unique S_2' subsite interactions are likely also involved (Kroger *et al.*, 2009). Contrary to RXP407 where the P_2 acetylated aspartate forms a salt bridge to R381 (PDB ID: 3NXQ), R381's side-chain is swivelled away from the active site in the 33RE-N domain structure (PDB ID: 4BXK) (Figure 2.1) (Douglas *et al.*, 2014). Although the unique N-domain Y369 does interact with the P_2 aminomethyl tetrazole moiety of 33RE, it is thus unclear whether it would display the same degree of N-selectivity as the parent compound RXP407.

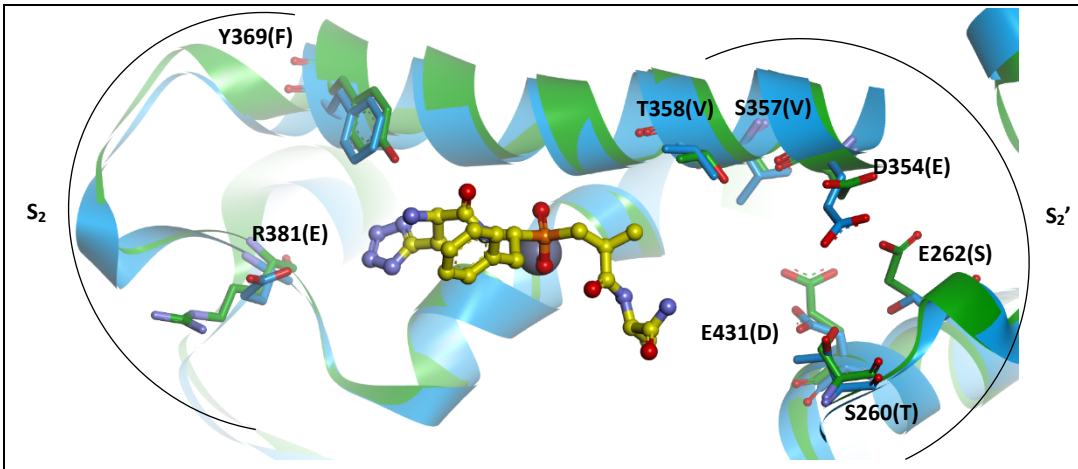


Figure 2.1: Structural alignment of the N- and C-domain illustrating direct binding of 33RE to unique S_2 residues but not distal, unique S_2' residues.
C-domain (PDB ID: 1O8A) is given in cyan, N-domain (PDB ID: 4BXK) in green, Zn as a grey sphere and the phosphinic inhibitor 33RE (co-crystallized with the N-domain) in yellow. Unique N-domain residues of the S_2 and S_2' subsites are labelled, and their C-domain counterparts given in parentheses.

The reason for the C-terminal or P_2' amide's importance in selectivity is also unclear since the N-domain RXP407/33RE co-crystal structures show that unique N-domain S_2' residues are distal from the inhibitor and thus the amide only interacts with S_2' residues conserved between the domains (Figure 2.1). Single or double ACE mutants with unique S_2' residues mutated from N- to C-domain counterparts also did not result in significant changes in RXP407 inhibitor binding (Kroger *et al.*, 2009). Yet, it has been shown that replacing the P_2' amide of RXP407 with a carboxylate leads to a complete lack of selectivity (Dive *et al.*, 1999).

While mutagenesis studies can aid in identification of specific residues and their role in ligand binding, they provide limited insight into the mechanism of inhibition at an atomic level. Conversely, crystal structures can be useful in providing mechanistic insight but are limited by factors such as crystal packing and crystallization pH. This artificial environment can potentially mask the effect that specific residue mutations have on binding, especially ones distal to the binding site. These factors are particularly important to consider when studying dynamic proteins with hinge-bending motions.

Based on homology modelling of ACE to its homologue ACE2 and normal mode analysis, it was previously suggested that the C-domain of ACE possesses a hinge mechanism for substrate binding (Watermeyer *et al.*, 2006). At present, all ACE N- and C-domain crystal structures are in a 'closed' conformation with a large, occluded central binding cleft formed by the two ellipsoid

subdomains. ACE2, however, has been crystallized in open (apo, PDB ID: 1R42) and closed (ligand-bound, PDB ID: 1R4L) conformations (Towler *et al.*, 2004). By aligning these structures to the N-domain of ACE it was observed that the unique N-domain S_2' residues reside in a region where structural changes occur upon transition to a closed conformation (Figure 2.2). These distal unique residues could thus indirectly mediate ligand binding by affecting protein hinging.

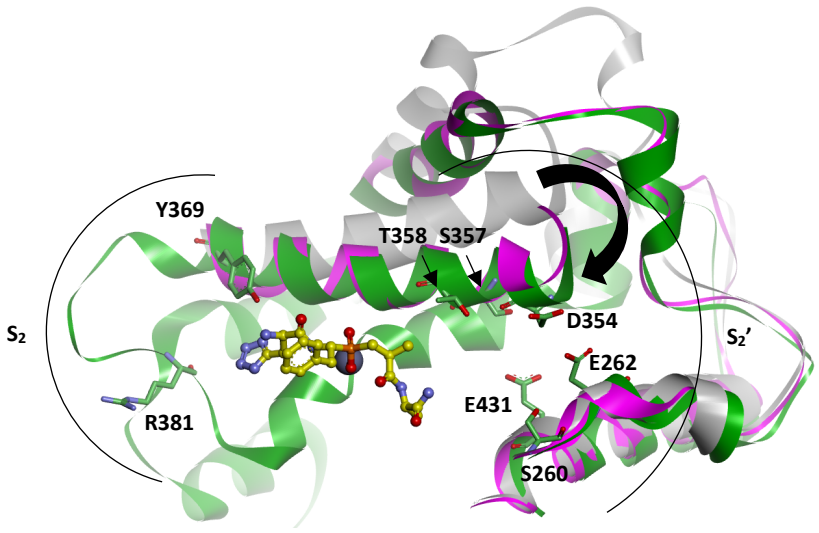


Figure 2.2: Structural alignment of the N-domain-33RE structure to the open and closed conformations of the ACE2 homologue.
N-domain (PDB ID: 4BXK) is given in green, ACE2 open conformation (PDB ID: 1R42) in grey and its closed conformation (PDB ID: 1R4L) in magenta. The black arrow indicates a ~10Å shift in an active site helix's position upon ligand binding to ACE2 and transition to the closed conformation. The unique N-domain S_2 and S_2' subsite residues are labelled, Zn shown as a grey sphere and the phosphinic inhibitor 33RE (co-crystallized with the N-domain) shown as yellow balls and sticks.

In this study, it is therefore proposed that the two ACE domains have different open conformations and hinge-bending motions and that this might influence inhibitor binding. This hypothesis is supported by three-dimensional electron microscopic reconstruction of negatively stained porcine sACE which revealed that the two domains' active site clefts have distinct shapes (Chen *et al.*, 2010). While the N-domain internal void area was circular, the C-domain interior space was noticeably larger and elongated.

2.2 Aims and objectives

The research described in this chapter aimed to investigate the contribution of all the unique N-domain residues in the S_2 and S_2' subsites towards N-selectivity using a combined *in vitro*/*in silico* approach. Provided that the selectivity predicted via docking could be confirmed *in vitro,* the mechanism of N-selectivity would be studied using the more drug-like molecule 33RE.

The objectives to achieve this aim were as follows:

1. Identify the active site residues crucial for selectivity by:
    a. Obtaining active site mutants with unique N-domain residues converted to their C-domain counterparts using site directed mutagenesis, cloning into an expression vector, mammalian cell transfection, heterologous protein expression and protein purification

    b. Evaluating the effect of mutation on inhibitor binding using an enzyme assay with quenched fluorogenic substrate to characterize inhibitor binding affinity

2. Elucidate the mechanism whereby the identified residues govern selectivity by:
    a. Creating a mutant having all N-domain S_2 and S_2' residues converted to their C-domain counterparts in a minimally glycosylated form amenable to crystallization using site-directed mutagenesis, sub-cloning, mammalian cell transfection, heterologous protein expression and protein purification

    b. Utilizing X-ray crystallography to solve the structure of this active site mutant in open (apo) and closed (inhibitor-bound) conformations for comparison to the N-domain active site cleft shape

    c. Assessing the effect of mutation on protein hinging using *in silico* mutagenesis and molecular dynamics simulations of apo and inhibitor-bound structures

    d. Monitoring the effect of ligand binding on wild-type or mutant protein structure *in vitro* using differential scanning fluorimetry

2.3 *In vitro* methodology

2.3.1 Site-directed mutagenesis

N- and C-domain ACE crystal structures (PDB IDs: 4BXK and 1O8A, respectively) were studied to identify the various unique residues thought to contribute to N-selectivity (Figure 2.1). Despite the modest changes observed upon single or double mutation of unique S_2' residues (Kroger *et al.*, 2009), the P_2' amide interaction of N-domain inhibitors is important for selective binding (Dive *et al.*, 1999). Since the N-domain residues differ in polarity and side-chain length from their C-domain counterparts, the S_2' subsite was progressively converted to C-domain (Table 2.1) to study the possibility of a synergistic environmental role in 33RE binding. The role of the S_2 subsite in N-selective binding was assessed using the previously generated and purified S_2 mutants Y369F, R381E and YR/FE (Kroger *et al.*, 2009).

Table 2.1: Active site mutants created via site-directed mutagenesis.

Active site mutant	N-domain residues	C-domain residues	Subsite
SE/VD	S357	V379	S_2'
	E431	D453	
SEDT/TSEV	S260	T282	
	E262	S284	
	D354	E376	
	T358	V380	
SEDSTE/TSEVVD	S260	T282	
	E262	S284	
	D354	E376	
	S357	V379	
	T358	V380	
	E431	D453	
Y369F [a]	Y369	F391	S_2
R381E [a]	R381	E403	
YR/FE [a]	Y369	F391	
	R381	E403	
SEDSTE_YR/TSEVVD_FE	S260	T282	S_2' / S_2
	E262	S284	
	D354	E376	
	S357	V379	
	T358	V380	
	E431	D453	
	Y369	F391	
	R381	E403	

[a]: mutants previously generated by Kröger *et al* (Kroger *et al.*, 2009). C-domain residues are given according to the standard tACE numbering (Ehlers *et al.*, 1989).

The S_2' mutants SE/VD, SEDT/TSEV and SEDSTE/TSEVVD were created as part of this thesis. The SE/VD mutant was chosen since single mutation of these S_2' residues had the greatest effect on RXP407 binding affinity (Kroger *et al.*, 2009) and the inverse SEDT/TSEV mutant used to study the combined role of the remaining unique S_2' residues. The possibility that synergism between all the unique residues could create a more favourable environment for P_2' amide binding was studied using the SEDSTE/TSEVVD mutant. In addition, these mutations were combined with the non-prime mutations to investigate inter-subsite synergism between the unique residues outlined in (Figure 2.1). The S_2'/S_2 mutant SEDSTE_YR/TSEVVD_FE was thus created.

Sequences encoding soluble forms of the fully glycosylated wild-type human N- and C-domain ACE proteins were previously cloned into the pBlueScript vector (pBS SKII+, Stratagene). The N-domain construct starts with an N-terminal signal sequence and terminates after residue D629 (Corradi *et al.*, 2006) whereas the C-domain construct starts with the testis ACE signal sequence, lacks the 36 unique N-terminal *O*-glycosylated residues and transmembrane region and terminates after S625 (Yu *et al.*, 1997). N-domain numbering is according to the N-terminal domain of somatic ACE minus the signal sequence. C-domain numbering in this chapter is according to the standard testis ACE minus the signal sequence. Structurally, residues 13 onwards of the N-domain and residues 40 onwards of the C-domain thus align (barring the N-domain 5 residue insertion) to yield a difference of 22 amino acids in the numbering of equivalent N- and C-domain residues, as shown in Acharya *et al* (Acharya *et al.*, 2003). These constructs thus represent the individual catalytic domains of somatic ACE.

Site directed mutagenesis was used to generate mutant N-domain constructs by converting the selected unique N-domain S_2 and/or S_2' subsite residues to their corresponding C-domain counterparts, as previously described (Kroger *et al.*, 2009). This was performed in an iterative manner to create constructs with different combinations of S_2 and/or S_2' mutations.

2.3.1.1 Design of mutagenesis primers

All site-directed mutagenesis primers used were 100% complementary and designed using the online WatCut tool (http://watcut.uwaterloo.ca/watcut/watcut/template.php). The E262S, E431D, Y369F and R381E primers were previously designed and available (Kroger *et al.*, 2009) whereas the S260T, D354E and V357S primers were synthesized by the Inqaba DNA Synthesis Unit (Inqaba Biotech Ltd.).

Each primer contained two sets of mismatches to introduce the desired amino acid change along with a silent mutation (Table A1, Appendix A1). This silent mutation created an additional restriction enzyme cleavage site to enable screening of resulting PCR amplicons for the codon change mutation. The pBS SKII+ S357V and pBS SKII+ S357V, T358V (ST/VV) constructs previously generated by Dr R.G. Douglas (Kroger *et al.*, 2009) were used as the initial PCR templates. The mutagenesis strategy is outlined in Figures 2.3-4.

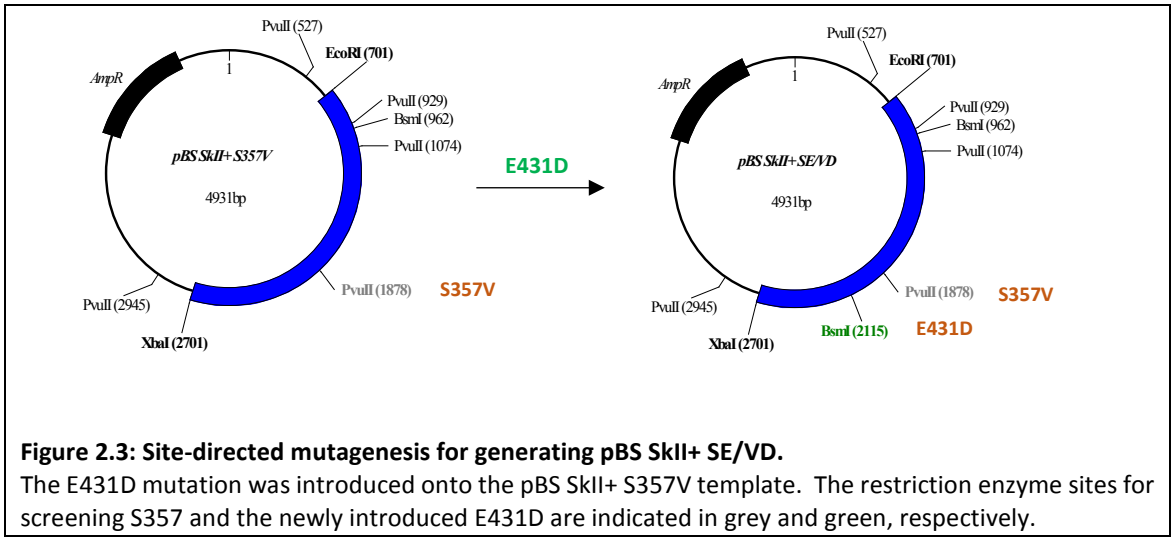


Figure 2.3: Site-directed mutagenesis for generating pBS SkII+ SE/VD.
The E431D mutation was introduced onto the pBS SkII+ S357V template. The restriction enzyme sites for screening S357 and the newly introduced E431D are indicated in grey and green, respectively.

The SEDT/TSEV mutant was generated by removing the S357V and E431D mutations from the pBS SkII+ SEDSTE/TSEVVD construct. A V357S primer was designed and PCR performed to reverse S357V (Figure 2.5). The E431D mutation was reversed to wild-type by sub-cloning of pBS SkII+ SEDTE/TSEVD and wild-type pcDNA 3.1+ N-domain with the enzymes *Eco*RI and *Bmg*BI.

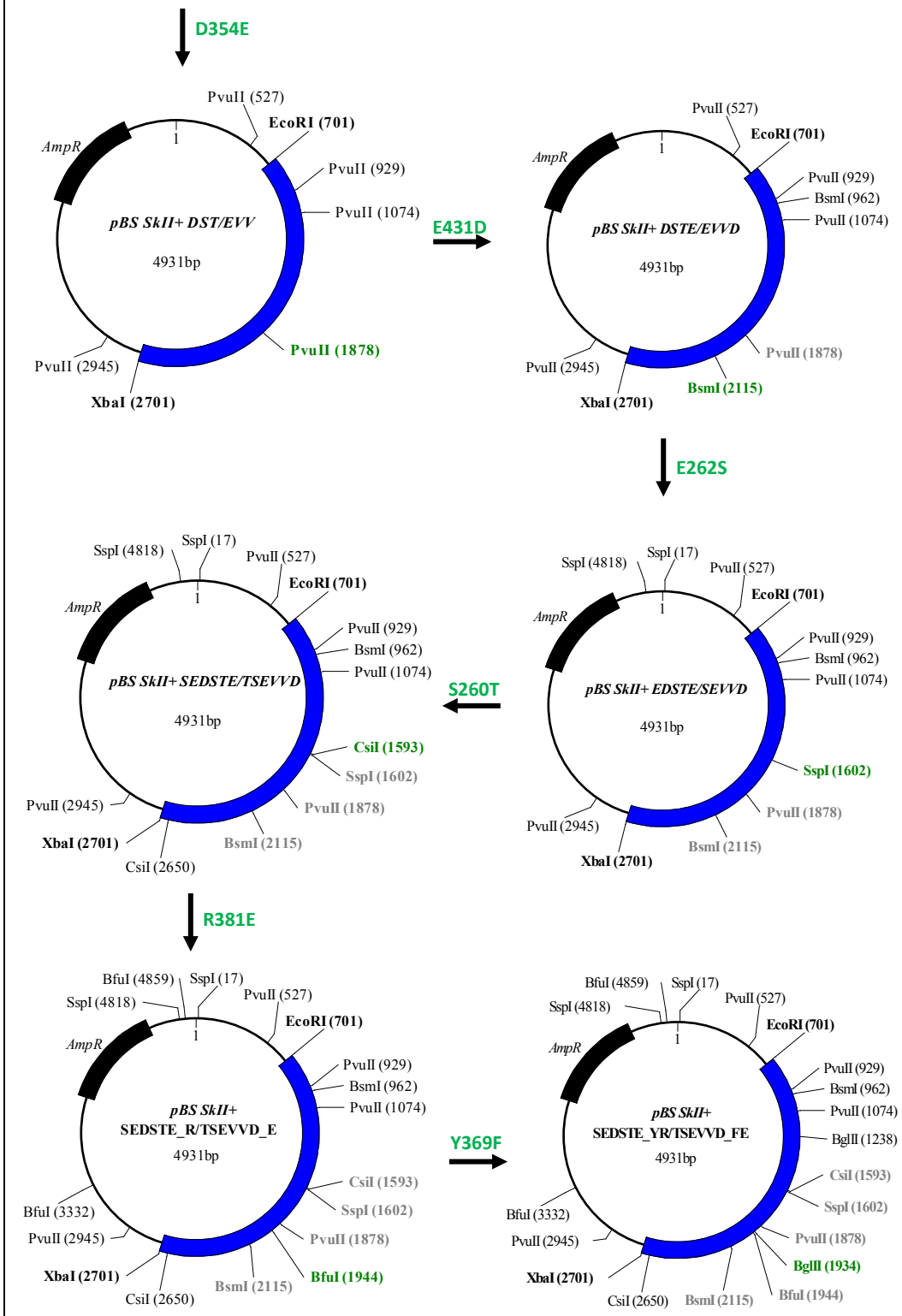


Figure 2.4: Sequential PCR strategy for generating pBS SkII+ SEDSTE/TSEVVD and pBS SkII+ SEDSTE_YR/TSEVVD_FE.

Mutations were iteratively added onto an initial pBS SKII+ ST/VV template. The newly introduced restriction enzyme screening sites are indicated in green on the vector maps for each stage of mutagenesis.

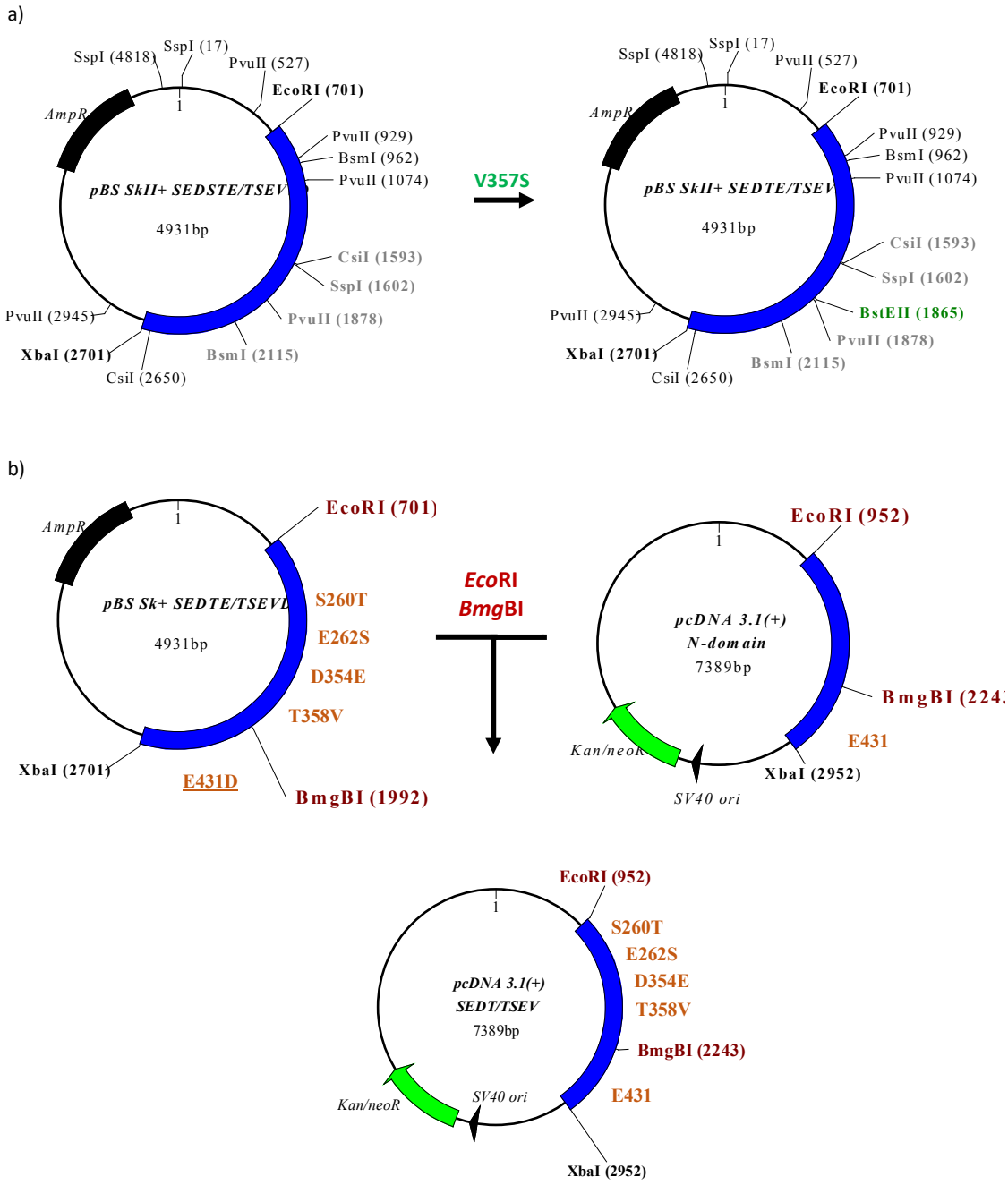


Figure 2.5: Strategy used to create the pcDNA 3.1+ SEDT/TSEV mutant.
Site-directed mutagenesis and sub-cloning were used to reverse S357V and E431D in pBS Sk+ SEDSTE/TSEVVD. The pBS SKII+ SEDSTE/TSEVVD construct was used as a template, a) the S357V mutation reversed by site-directed mutagenesis and b) the E431D mutation reversed by sub-cloning of pBS SkII+ SEDTE/TSEVD and pcDNA 3.1+ N-domain with *Eco*RI and *Bmg*BI. The introduced V357S site is indicated in green on the vector map.

2.3.1.2 Site directed mutagenesis procedure

A *Dpn*I site-directed mutagenesis approach was followed to create mutant N-domain constructs. Final concentrations per PCR reaction were 0.3µM forward and reverse primers (Inqaba Biotech, Ltd.), 0.3µM dNTPs (Thermo Scientific dNTP Mix), a total mass of 100ng template DNA and one unit of KAPA HiFi™ high fidelity polymerase (KapaBiosystems Ltd., SA). Reactions for each mutagenesis run were carried out at 0mM and 2mM Mg^{2+} concentrations to determine the optimal conditions for primer specificity and polymerase activity using $MgCl_2$ (KapaBiosystems Ltd., SA). PCR reactions were performed in a Thermal cycler (Bio-Rad MyCycler™) with parameters as outlined in Table 2.2.

Table 2.2: PCR cycling parameters used for site-directed mutagenesis.

Step	Temperature (°C)	Time	Number of cycles
Initial Denaturation	94	2 min	1
Denaturation	98	20 secs	25
Primer Annealing	85	30 secs	
Extension	68	2.5 min	
Final Extension	68	5 min	1
Cooling	4	∞	

The resulting amplicons underwent agarose gel electrophoresis (protocols and recipes in Appendix A5) and visualization with Ethidium Bromide (EtBr) staining under ultraviolet (UV) light. PCR products from mutagenesis runs showing the highest amplification and specificity were selected and digested at 37˚C for 18 hours with methylation-specific restriction enzyme *Dpn*I (Fermentas Life Sciences) to remove remaining template DNA. Subsequently, 20µl of the digested amplicons were transformed into chemically competent DH5α *Escherichia coli* cells (protocols in Appendix A3). Transformed cells were plated onto Luria agar plates with Ampicillin selection (100µg/ml (Sigma-Aldrich® Co.)) and incubated for 18 hours at 37˚C (Appendix A2). Resistant colonies were picked and cultured overnight in Ampicillin-containing Luria Bertani (LB) medium from which the DNA was then crudely isolated using a STET prep (protocols and recipes in Appendix A2 and A4).

2.3.2 Confirmation of introduced mutations

DNA isolated from cultured clones was digested with the appropriate screening enzymes as outlined in Table A1 (Appendix A1, (*Bfu*I, *Bgl*II, *Bsm*I, *Bst*EII, *Csi*I and *Ssp*I from Fermentas Ltd., *Pvu*II from New England Biolabs Inc.)). The DNA fragments produced were resolved by agarose gel electrophoresis. Positive incorporation of the desired mutations was identified by differences in the mutants' DNA banding pattern compared to the template DNA due to simultaneous introduction of restriction enzyme sites.

High quality plasmid DNA was isolated from colonies positive for the desired mutations using the GeneJET™ Plasmid Miniprep kit (Thermo Fisher Scientific Inc.). The restriction enzyme digest results were subsequently confirmed and absence of spurious mutations verified by nucleotide sequencing. Internal N-domain primers (Inqaba Biotech, Ltd.) and T3 and T7 vector primers (Promega Corp.) flanking the N-domain were used for sequencing of the entire N-domain coding region (Gordon, 2011). Samples were sent to the Stellenbosch University Central Analytical Facility for sequencing.

2.3.3 Cloning of mutant constructs into a mammalian expression vector

Positive mutant constructs were cloned from the pBS SkII+ bacterial sequencing vector into the pcDNA 3.1+ mammalian expression vector (Invitrogen™) to enable transfection into mammalian cells and optimal expression, glycosylation and folding of the proteins. The cloning strategy involved a double digest of 600ng mutant pBS SkII+ construct with one unit of *Eco*RI and a two-fold excess of *Xba*I (both restriction enzymes from Fermentas Ltd.). The wild-type N-domain pcDNA 3.1+ vector was similarly digested to allow compatible ends for cloning of the mutant N-domain insert into the expression vector. This approach was followed for the SE/VD, SEDSTE/TSEVVD and SEDSTE_YR/TSEVVD_FE mutants. For the SEDT/TSEV mutant, 600ng of the pBS SkII+ SEDTE/TSEVD construct (produced by site-directed mutagenesis) was digested with one unit of *Bmg*BI and a 2-fold excess of *Eco*RI. This strategy allowed simultaneous removal of the E431D mutation and cloning into pcDNA 3.1+ to yield pcDNA 3.1+ SEDT/TSEV. Following restriction enzyme digests, the DNA fragments were separated by agarose gel electrophoresis and visualized with EtBr staining under UV light. Insert and vector bands corresponding to 2000bp and 5389bp, respectively, (for all mutants except SEDT/TSEV which had 1291bp and 6098bp, respectively) were excised and the GeneJET™ Gel Extraction kit (Thermo Fisher Scientific Inc.) used for DNA purification. Ligation was performed by combining vector and insert

DNA at a 1:3 ratio in a reaction mix containing T4 DNA Ligase buffer and 5 units of T4 DNA Ligase (Fermentas Ltd.). After incubation at room temperature for two hours, chemically competent DH5α cells were transformed with the ligation reaction, plated and grown for 18 hours under Ampicillin selection (protocols and recipes in Appendix A2-4). Resistant clones were picked, grown in Ampicillin-containing LB medium and the DNA isolated as before using a STET prep (protocols and recipes in Appendix A2-A4).

Successful cloning of the mutated inserts was confirmed by double digests with the cloning enzymes, *Eco*RI and *Xba*I, in addition to single digests with screening enzymes outlined in Table A1 (Appendix A1). Agarose gel electrophoresis allowed identification of positive clones from which DNA was then bulk purified using the Zyppy™ Plasmid Midiprep kit (The Epigenetics Company).

2.3.4 Generation of a minimally glycosylated SEDSTE_YR/TSEVVD_FE mutant

The N-domain is routinely crystallized in a minimally glycosylated form where six of the nine PNGs have been mutated to glutamine residues. The resulting construct only retains glycosylation at sites n3, n8 and n9 (Anthony *et al.*, 2010). To enable crystallization of the SEDSTE_YR/ TSEVVD_FE active site mutant, the asparagine residues at sites n1, n2, n4, n5, n6 and n7 were converted to glutamine residues.

Sites n1, n2, n4, n5 and n6 were situated upstream of the active site mutations. They were thus removed by sequentially digesting 600ng of the fully glycosylated pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE construct and an N-domain construct only containing site n3 (pcDNA 3.1+ N-domain n3 (Redelinghuys, 2006)) with Bsu*36I* in 1x Tango buffer for 1 hour at 37°C and then Eco*RI* in 2x Tango buffer for 1 hour at 37°C. The resulting bands were resolved on a 1% agarose gel and DNA purification, ligation, transformation and DNA isolation performed as described in section 2.3.3. A double digest of *Eco*RI and *Xba*I was performed as before to screen for the presence of a 2000bp coding region and thus generation of the pcDNA 3.1+ SEDSTE_YR/ TSEVVD_FE n3789 construct. Site n7 was removed via site directed mutagenesis using the primer previously designed by Dr P. Redelinghuys (Redelinghuys, 2006). The PCR parameters used were as described in section 2.3.1.2. Positive introduction of the desired mutation, and thus generation of the pcDNA 3.1+ SEDSTE_YR/ TSEVVD_FE n389 construct, was confirmed by restriction enzyme screening for the simultaneously introduced *Nde*I site. Nucleotide

sequencing was performed to confirm the presence of all active site mutations and N>Q mutation of sites n1, n2, n4, n5, n6 and n7.

2.3.5 Protein expression

The Chinese Hamster Ovary cell line (CHO-K1) was used for ACE protein expression. Cells which were previously stably transfected with wild-type N- or C-domain (Kroger *et al.*, 2009) were reconstituted into growth medium. They were grown to confluency in T-175cm^2 flasks (Nunc Ltd.) before replacing the growth medium with harvesting medium (constituents in Appendix A6). Culture medium was harvested every 48 hours and the pooled harvests stored at -20°C until protein purification.

To express the mutant proteins, untransfected cells were plated in 100mm^2 plates (Nunc Ltd.) and incubated overnight at 37°C, 5% CO_2 and 80% humidity in complete growth medium. Three hours prior to transfection, fresh complete growth medium was applied. At 40% confluency, the cells were transfected with mutant N-domain pcDNA 3.1+ constructs using the calcium phosphate Profection® Mammalian Transfection System (Promega Corp.). A solution of 10µg DNA in $CaCl_2$ was bubbled through a HEPES Buffered Saline (HBS) solution to form a fine DNA precipitate. This was added to the cells in a dropwise manner with swirling to allow even distribution across the plate. Following 4 hour's incubation, a 2-minute glycerol shock (15% glycerol in 1X PBS, pH7.5) was performed to increase the transfection efficiency before removing the DNA solution. Cells were incubated overnight at the aforementioned conditions in fresh complete growth medium until confluency was reached.

After 24 hours, selection of positively transfected cells was commenced by replacing the growth medium with complete growth medium containing 0.9mg/ml of the mammalian antibiotic, G418 disulphide salt (Sigma-Aldrich® Co.). Resistant clones were picked into 48-well plates containing growth medium and kept under G418 selection until confluency was reached. Growth medium was subsequently replaced with harvesting medium and the cells incubated for 48 hours. After incubation, a fluorometric assay was performed to identify clones expressing high levels of ACE (see section 2.3.6). The highest-expressing clone was selected for each mutant, grown to confluency in T-25cm^2 flasks (Nunc Ltd.) and subsequently split into T-150cm^2 flasks. Upon reaching confluency, harvesting medium was applied to the T-150cm^2 flasks and incubated for

48 hours to allow protein expression. Medium was harvested, pooled and stored at -20˚C until protein purification.

2.3.6 End-point ACE fluorogenic assay

ACE activity was assessed during protein expression and purification using the non-domain selective fluorogenic substrate Z-phenylalanylhistidylleucine (Z-FHL) (Bachem Ltd.) (Schwager *et al.*, 2006). Two-hundred and twenty milligrams of Z-FHL substrate was dissolved in 2ml NaOH (0.28M) before addition of 2ml distilled water in a dropwise manner with vortexing. A 20mM stock solution was obtained by addition of a further 16ml distilled water and 1mM working solutions subsequently prepared in assay buffer (100mM K_2HPO_4/KH_2PO_4, pH 8.3, 300mM NaCl, 10µM $ZnSO_4$).

Six microliters of harvested medium or 3µl of pure protein was added to 30µl of 1mM Z-FHL in triplicate wells of a 96-well plate. The plate was incubated at 37°C on an orbital shaker for 15 minutes to allow substrate hydrolysis by ACE. This was followed by termination of the reaction with 177µl 0.28M NaOH and addition of 16µl O-phthaldialdehyde (20mg/ml (Sigma-Aldrich® Co.)) in methanol. Incubation for 10 minutes at room temperature with shaking allowed derivatization of the HL sequestered by Z-FHL hydrolysis. The reaction was stopped by addition of 30µl HCl (3M) and the fluorescence intensity detected at excitation and emission wavelengths of 360nm and 485nm, respectively, using a fluorescence spectrophotometer (Cary Eclipse Varian Inc.). A standard curve was constructed using His-Leu (HL) standards (Sigma-Aldrich® Co.) prepared in assay buffer to relate fluorescence intensity to moles of HL, and thus enzyme activity (Appendix A7).

2.3.7 Protein purification

The purified Y369F, R381E and YR/FE proteins were a kind gift from Dr R. G. Douglas (Kroger *et al.*, 2009). ACE proteins harvested from CHO-K1 cells were purified using lisinopril-sepharose affinity chromatography, as previously described (Bull *et al.*, 1985). This involves binding of ACE in the medium to the potent inhibitor, lisinopril, coupled to sepharose beads. To increase the affinity for lisinopril binding, a final concentration of 800mM NaCl was added to N-domain medium (Deddish *et al.*, 1994). The column was prepared and equilibrated by washing with wash buffer (20mM HEPES pH 7.5, 500mM NaCl) for 15 minutes. Pooled harvested medium was

loaded onto the column and wash buffer applied for 12 hours to remove any remaining medium components from the column. This was followed by dissociation of the protein with elution buffer (50mM Boric acid, pH 9.5). Fractions were collected from the first point of elution as monitored by a UV detector (Settings: 5mV, 100m/h, A_{280nm}). Activity of each fraction was determined with the described Z-FHL assay (section 2.3.6) and the most active fractions pooled into SnakeSkin® dialysis tubing (Pierce Biotechnology Ltd.). Pooled protein solutions were dialyzed for 24 hours at 4˚C in dialysis solution 1 (5mM HEPES pH 7.5, 0.1mM PMSF in ethanol) after which they were transferred into dialysis solution 2 (50mM HEPES pH 7.5, 0.1mM PMSF in ethanol) for a further 24-hour period at 4˚C.

Purity of the resulting proteins was assessed by means of sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). A 10% acrylamide gel was used to separate proteins based on their molecular mass before staining the gel with Coomassie dye (Appendix A8) for visualization. The concentration of pure protein was determined from the absorbance (extinction coefficient of 162 070 M^{-1} cm^{-1} for N-domain and 143 620 M^{-1} cm^{-1} for C-domain; molecular mass of 72.36 kDa for N-domain and 77.82 kDa for C-domain) with the use of a NanoDrop® spectrophotometer.

2.3.8 Preparation of substrate and inhibitor for kinetic assays

The non-selective, quenched fluorogenic ACE substrate (Abz)-FRK-(Dnp)-P-OH was kindly donated by Dr A.K. Carmona (Universidad Federal de São Paulo, Brazil). The peptide was dissolved in 100% DMSO and the concentration determined by measuring the absorbance at 365nm and extinction coefficient of 17300 M^{-1} cm^{-1}. Subsequent dilutions were made in assay buffer (50mM HEPES, pH 6.8, 10µM $ZnCl_2$, 200mM NaCl).

The phosphinic inhibitor 33RE was synthesized by Dr R.K. Sharma (Department of Chemistry, University of Cape Town, South Africa). Lyophilized 33RE was dissolved in distilled water to yield a stock concentration of 10mM. Subsequent dilutions were made in an assay buffer consisting of 50mM HEPES pH 6.8, 200mM NaCl and 10µM $ZnCl_2$.

2.3.9 Preparation of an Abz standard curve

The unquenched fluorescent group o-Aminobenzoic acid (Abz) coupled to glycine (Abz-Gly, Bachem Ltd.) was used to correlate the fluorescence intensity to the amount of Abz present. A 1mM stock of Abz-Gly was prepared in distilled water and serially diluted in assay buffer (50mM HEPES pH 6.8, 10µM $ZnCl_2$ and 200mM NaCl). Three hundred microliters of Abz-Gly at each concentration was added to triplicate wells of a 96-well plate and the fluorescence intensities read at excitation and emission wavelengths of 320nm and 420nm, respectively, using a fluorescence spectrophotometer (Cary Eclipse Varian Inc.). The fluorescence intensity was plotted against picomoles of Abz-Gly and a standard curve obtained (Appendix A9) by linear regression analysis (GraphPad Prism® v6.01).

2.3.10 Catalytic efficiency

The amount of active protein was determined immediately after purification by calculating the specific activity from Z-FHL (Schwager *et al.*, 2006) and Bradford (Bradford, 1976) assays. Binding and hydrolysis of the fluorogenic Abz-FRK(Dnp)P-OH substrate was determined as previously described in an assay buffer consisting of 50mM HEPES (pH 6.8), 200mM NaCl and 10µM $ZnCl_2$ (Carmona *et al.*, 2006). Briefly, a 50µM solution of Abz-FRK(Dnp)P-OH was prepared in assay buffer (50mM HEPES, pH 6.8, 10µM $ZnCl_2$, 200mM NaCl) and variable volumes added in triplicate to a 96-well plate on ice. A constant volume of enzyme was added to all wells and the final volume adjusted to 300µl with buffer to give a constant amount of enzyme with a range of substrate concentrations. The hydrolysis of Abz-FRK(Dnp)P-OH from triplicate wells were continuously monitored at excitation and emission wavelengths of 320nm and 420nm, respectively, using a fluorescence spectrophotometer (Cary Eclipse, Varian Inc.).

Since quenching of the released Abz fluorescence signal occurs at high Abz-FRK(Dnp)P-OH concentrations, a phenomenon known as the inner filter effect, initial velocities were adjusted by constructing a correction curve of Abz-Gly and (Abz)-YRK-(Dnp)-P-OH (Appendix A10) as previously described (Carmona *et al.*, 2006; Liu *et al.*, 1999).

Initial reaction velocities were determined from the fluorescence intensity at each Abz-FRK(Dnp)P-OH concentration using the Abz-Gly standard curve (Appendix A9), adjusted using the second order polynomial of the correction curve (Appendix A10) and plotted against

substrate concentration to obtain the K_m and k_{cat} kinetic parameters using the Michaelis-Menten model (GraphPad Prism® v 6.01). Catalytic efficiency of the Y369F, R381E and YR/FE mutants were previously determined and described in Kröger *et al* (Kroger *et al.*, 2009).

2.3.11 Inhibition kinetics

Inhibition assays were performed in triplicate wells of two independent experiments for wild-type N-domain, C-domain and mutant proteins to determine the dissociation constant for binding of inhibitor to enzyme (K_i). The Dixon method (Dixon, 1953) was used which involved measurement of initial reaction velocities at constant enzyme but two different substrate concentrations and a range of inhibitor concentrations. Dixon plots were determined for wild-type as well as mutant enzymes to study the effect of active site mutations on inhibitor binding and thus selectivity.

An equal volume of enzyme (40nM) was incubated for 5 minutes with an appropriate concentration range of phosphinic inhibitor at ambient temperature. Twenty microliters of the enzyme-inhibitor mix was aliquoted in triplicate to a 96-well plate followed by 280µl of (Abz)-FRK(Dnp)P-OH substrate. To construct a Dixon plot, two parallel assays with different substrate concentrations (4µM and 8µM) were performed for each enzyme. Fluorescence due to liberation of the fluorescent Abz group from the quenching 2,4-dinitrophenyl (Dnp) group was detected continuously for 30 minutes at excitation and emission wavelengths of 320nm and 420nm, respectively, using a fluorescence spectrophotometer (Cary Eclipse Varian Inc.). Since detected fluorescence is directly proportional to the amount of Abz released, and thus reaction velocity, the inverse fluorescence was plotted as a function of inhibitor concentration at two different substrate concentrations to yield a Dixon plot (Appendix A12).

2.4 *In silico* methodology

2.4.1 Protein structure preparation

Files containing the atomic coordinates and structure factors for the N-domain-33RE co-crystal structure were obtained from the protein data bank (PDB, www.rcsb.org; accession code 4BXK). Before using the deposited structure for modelling and molecular simulations the model's fit to the electron density map was validated. Phenix ReadySet was used to add hydrogens to the

protein and ligand prior to iteratively refining the structure in Phenix Refine with optimization for X-ray weight and stereochemistry (Adams *et al.*, 2010). Coot was used in conjunction with Phenix for visual inspection of electron density and adjustment of side chain rotamers/chi angles, if necessary (Emsley *et al.*, 2010).

Histidine tautomers and amide orientations (asparagine and glutamine) were visually inspected and altered to ensure optimal hydrogen bonding using the Schrödinger Suite's Protein Preparation Wizard (Sastry *et al.*, 2013). Optimization was performed on protein chain A in the presence of crystallographic waters but the absence of protein chain B and crystallographic sugars. The crystallographic waters were deleted after successful protein optimization.

All subsequent simulations were performed using Biovia Discovery Studio v4.0 (Biovia, 2013). Valencies of the ligand structure were confirmed and the ionized phosphonate's negative charge delocalized over the OAI, OAH and PBG atoms. The 33RE P_2' carboxylate analogue was created by replacing the P_2' amide with a carboxylate having a partial double bond to each oxygen. Even though protein chain A was more complete than B, it still had poor electron density for residues L129 - T133. The missing loop of residues P130, Q131 and K132 was built using the random tweak algorithm in Biovia Discovery Studio (Shenkin *et al.*, 1987). All atoms except residues 129-133 were held fixed in position while further optimizing the loop's geometry through 1000 steps of steepest descent minimization (rms gradient = 3) and conjugate gradient minimization (rms gradient = 0.01). The protein and peptidomimetic ligand were protonated at pH 7.4 and typed with the CHARMm forcefield using the Momany-Rone partial charge method (Momany and Rone, 1992). This forcefield allowed typing of the ligand tetrazole and peptidomimetic moieties. To our knowledge, no parameters have yet been published for zinc coordination to phosphonate atoms. To avoid estimation of these parameters but still maintain the catalytic zinc ion, distance restraints (force constant 50kcal/(mol.Å^2)) were applied between the zinc (+2 charge) and its coordinating atoms: the ligand phosphonate OAI and OAH, H361 NE2, H365 NE2 and E389 OE1.

The apo structure was prepared by deleting the ligand and only applying distance restraints between zinc and H361 NE2, H365 NE2 and E389 OE1. The mutant structures were all prepared from these two typed wild-type structures. With each mutation or deletion of ligand, optimal

hydrogen bonding within a 5Å radius of the change was verified by visual inspection using the Schrödinger Suite's Protein Preparation Wizard (Madhavi Sastry *et al.*, 2013).

2.4.2 Molecular dynamics

Molecular dynamics simulations were performed on a total of six 33RE-bound, three 33RE P_2' carboxylate analogue-bound and six apo structures.

i. Solvation and minimization:

Fixed atom constraints were applied to all protein and ligand atoms before explicitly solvating the systems. The water molecules were added using the explicit periodic boundary TIP3P model (Jorgensen *et al.*, 1983) with a truncated octahedron cell shape and a minimum distance of 10Å between protein atoms and the solvent boundary. This cell shape was the best fit for the globular ACE molecule and comparable to literature (Papakyriakou *et al.*, 2007).

Counterions to a concentration of 0.145M NaCl were added to neutralize the system and mimic physiological and *in vitro* assay conditions. The solvent box was subjected to one hundred steps of steepest descent minimization (RMS gradient = 3) and subsequently 2500 steps of conjugate gradient minimization (RMS gradient = 0.05). Thereafter, the fixed atom constraints were released and the entire system minimized with 1500 steps of conjugate gradient minimization (RMS gradient 0.05) while harmonically constraining the protein backbone (force constant = 10kcal/(mol.$Å^2$)). Long-range electrostatics were calculated using the Particle Mesh Ewald method (Darden *et al.*, 1993) with an interpolation order of 4 and spacing of 0.34 Å.

ii. Heating:

The minimized structures were heated for 2500 steps from the initial temperature of 50K to the target 300K by adjusting the velocity every 50 steps with a 2fs time step.

iii. Equilibration:

To ensure proper distribution of energy, the systems were equilibrated at the target temperature with constant pressure (1atm) for 55000 steps. The Nosé-Hoover thermostat (Nosé, 1984; Hoover, 1985) was used for temperature control and the Langevin piston (Feller *et*

al., 1995) for pressure control with a piston mass of 1000 AMU and collision frequency of 25ps^{-1}. Velocities were monitored and adjusted every 50 steps if necessary. The Leapfrog-Verlet dynamics integrator was used with a 2fs timestep to numerically integrate the equations of motion.

iv. Production:

Production runs of 12.55ns were carried out at physiological conditions in the NPT ensemble (1atm, 300K) with all restraints (except the Zn^{2+}-binding distance restraints) having been released. Snapshots of the trajectories were taken every 5ps for a total of 2510 frames per structure. The non-bond list radius was set to 12Å with higher and lower cut-offs of 10Å and 8Å, respectively, for all simulations. The SHAKE constraint, which fixes all bonds involving hydrogens, was applied to all runs except the explicit solvent box minimizations.

2.4.3 Trajectory analysis

Prior to analysis, the trajectory files were processed to remove all water molecules, sodium and chloride ions. The sampling was increased to 10ps intervals to reduce computational cost and enable analysis of the fifteen structures. Energy and temperature of the system were monitored during the equilibration and production stages to ensure stability and equilibrium. Trajectory snapshots were analysed for backbone root mean square deviation (RMSD) against the first equilibrated frame with backbone fitting to remove the translational and rotational motions of the protein molecule over time. The fitted conformations were subsequently analysed for amino acid root mean square fluctuation (RMSF), protein radius of gyration, ligand poses and per residue ligand interaction energy decomposition (within 5Å radius of ligand) using Biovia Discovery Studio v4.0 (Biovia, 2013).

Ligand stability in the active site was visualized by dividing the 955 frames into 100 frame sections (equal to 1ns) to extract ten average conformations for each structure. The effect of mutation on active site size over time was evaluated by computing the volume of the central cavity for the 10 average conformations per structure using the online server CASTp (Binkowski *et al.*, 2003). The 33RE-bound structures' volumes were compared to their respective apo forms using a two-tailed, paired, parametric t-test with a 99% confidence level (GraphPad Prism® v6.01).

Protein backbone motions and subdomain hinging was studied by RMSD clustering, covariance and principal component analyses using Biovia Discovery Studio v4.0 (Biovia, 2013) and the Bio3D package in R (Grant *et al.*, 2006). Clusters of conformations were identified by calculating the RMSD with backbone fitting over all frames across the simulation using the hierarchical Unweighted Pair Group Method with Arithmetic mean (UPGMA). Correlated motions of backbone atoms were extracted and visualized on covariance heatmaps. Diagonalization of the covariance matrix through principal component analysis (PCA) removed the high frequency protein motions and allowed visualization of long-range, low frequency backbone motions conserved throughout the simulations. The resulting eigenvectors and eigenvalues were visualized as porcupine plots using a python script written by Robert Schulz in VMD (Humphrey *et al.*, 1996). Images were made with Biovia Discovery Studio v4.0 (Biovia, 2013), UCSF Chimera v1.10.1 (Pettersen *et al.*, 2004), and VMD (Humphrey *et al.*, 1996) software. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

2.5 *In vitro* methodology

2.5.1 Differential scanning fluorimetry

Protein melting temperatures and thermal stabilization by 33RE were determined using the Life technologies™ Protein Thermal Shift assay according to manufacturer's instructions. Briefly, pure protein and 33RE were incubated in assay buffer (50mM HEPES pH6.8, 200mM NaCl and 10µM $ZnCl_2$) at final concentrations of 1.3µM and 0-85µM, respectively, for 10 minutes at ambient temperature. The thermal shift fluorogenic dye was added to protein-inhibitor reactions in triplicate wells of 96-well RT-PCR plates on ice and the wells sealed with StarSeal® optically clear film. Fluorescent signal due to protein thermal denaturation was monitored over a temperature range of 25-90°C at 0.5°C intervals using a BioRad® CFX-6 RT-PCR machine. Two independent experiments were performed for each protein. Melting temperatures (T_m) were determined from the peak of the obtained first derivative results and averaged across the triplicate wells. The degree of thermal stabilization (ΔT_m) calculated as the difference in average T_m in the presence vs absence of 33RE was plotted against concentration and fitted to a one-site binding model (GraphPad Prism® v6.01). The maximum stabilization was calculated from these curves for each protein and that of the wild-type compared to the mutants' using an unpaired, two-tailed parametric t-test with a 99% confidence level (GraphPad Prism® v6.01).

2.6 Results and discussion

2.6.1 Generation of N-domain active site mutant constructs

The N-domain construct starts with an N-terminal signal sequence and terminates after residue D629 (Corradi *et al.*, 2006) whereas the C-domain construct starts with the testis ACE signal sequence, lacks the 36 unique N-terminal *O*-glycosylated residues and transmembrane region and terminates after S625 (Yu *et al.*, 1997). N-domain numbering is according to the N-terminal domain of sACE minus the signal sequence. C-domain numbering in this chapter is according to tACE minus the signal sequence but including the 36 unique N-terminal residues. The first amino acid present in the C-domain construct used here is thus numbered as 37. Site-directed mutagenesis yielded positive amplification of all the desired constructs (Table A1) at magnesium concentrations of 0mM and 2mM, 100ng template DNA and primer annealing temperature of 85˚C. The restriction enzymes screens performed with the intermediate pBS SKII+ vectors are not shown but positive incorporation of the desired mutations was observed for all constructs. Nucleotide sequencing confirmed this result and showed no spurious introductions.

The SEDT/TSEV mutant was generated from pBS SKII+ SEDSTE/TSEVVD by V357S site-directed mutagenesis and sub-cloning to replace E431D by wild-type. Sequencing as well as digests of template and amplicon constructs with *Bst*EII and *Eco*RI (data not shown) confirmed positive introduction of the V357S mutation and thus generation of pBS SkII+ SEDTE/TSEVD. The E431D mutation present in pBS SkII+ SEDTE/TSEVD was reverted by replacing the C-terminal coding region with wild-type by means of an *Eco*RI/*Bmg*BI digest. The resulting pcDNA 3.1+ SEDT/TSEV was digested with all the restriction enzymes for which sites were introduced via PCR. Differences in banding patterns compared to wild-type confirmed the presence of all desired mutations and positive removal of S357V and E431D (Figure 2.6). This result was confirmed by nucleotide sequencing.

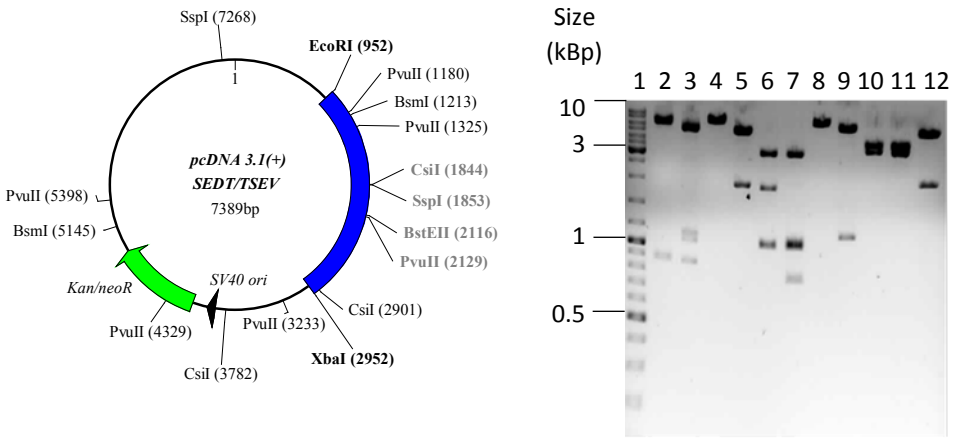


Figure 2.6: Restriction enzyme screen for presence of all pcDNA 3.1+ SEDT/TSEV mutations.
Vector map illustrates introduced screening sites in grey. Right hand panels show electrophoresis results for the respective restriction digests (lane 1: O'GeneRuler^{TM} DNA Ladder Mix (Fermentas), lanes 2, 4, 6, 8, 10: pcDNA 3.1+ N-domain, lanes 3, 5, 7, 9, 11, 12: pcDNA 3.1+ SEDT/TSEV. Lanes 2-3: *Csi*I, lanes4-5: *Ssp*I, lanes 6-7: *Pvu*II, lanes 8-9: *Eco*RI/*Bst*EII, lanes 10-11: *Bsm*I and lane 12: *Eco*RI/*Xba*I).

The prepared SE, SEDSTE and SEDSTE_YR coding regions were cloned into the mammalian expression vector by digesting the pBS SKII+ mutant and pcDNA 3.1+ N-domain constructs with *Eco*RI and *Xba*I (SE/VD shown as a representative example in Figure 2.7). Subsequent ligation of the 2000bp mutant insert with the 5389bp pcDNA 3.1+ vector yielded 7389bp plasmids encoding the respective mutants.

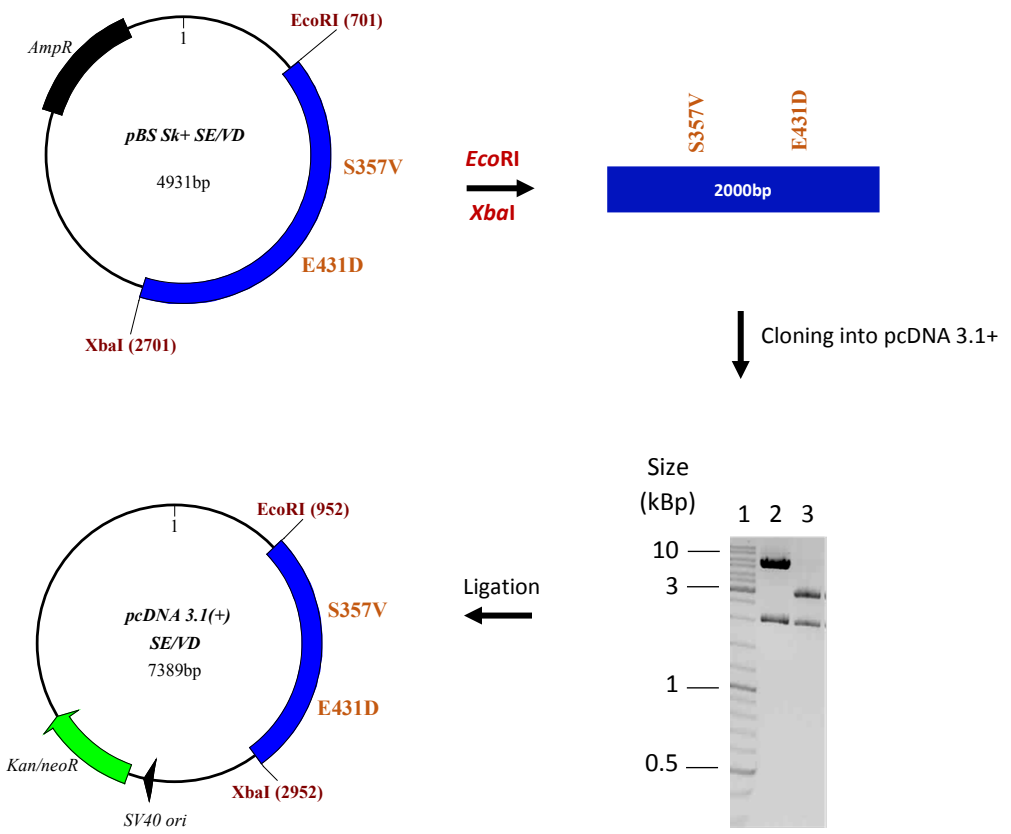


Figure 2.7: Cloning of N-domain mutant SE/VD into mammalian pcDNA 3.1+ expression vector.
a) Cloning strategy with restriction enzyme sites introduced during mutagenesis shown in grey. The *Pvu*II site at 1878bp was previously introduced with the S357V mutagenesis. b) Restriction enzyme digests (1% agarose gel) for cloning of the SE/VD insert into pcDNA 3.1+. Lane 1: O'GeneRuler™ DNA Ladder Mix (Fermentas), lane 2: pcDNA 3.1+ N-domain digested with *Eco*RI & *Xba*I, lane 3: pBS SkII+ SE/VD digested with *Eco*RI & *Xba*I. pcDNA 3.1+ vector and N-domain insert sizes correspond to 5389bp and 2000bp, respectively.

All the resulting pcDNA 3.1+ constructs were screened with the restriction enzymes for which sites were introduced via PCR. Differences in banding patterns compared to wild-type confirmed the presence of all desired mutations for the SE/VD (Figure 2.8), SEDSTE/TSEVVD (Figure 2.9) and SEDSTE_YR/TSEVVD_FE (Figure 2.10) constructs. These results were verified by nucleotide sequencing and the absence of spurious mutations confirmed.

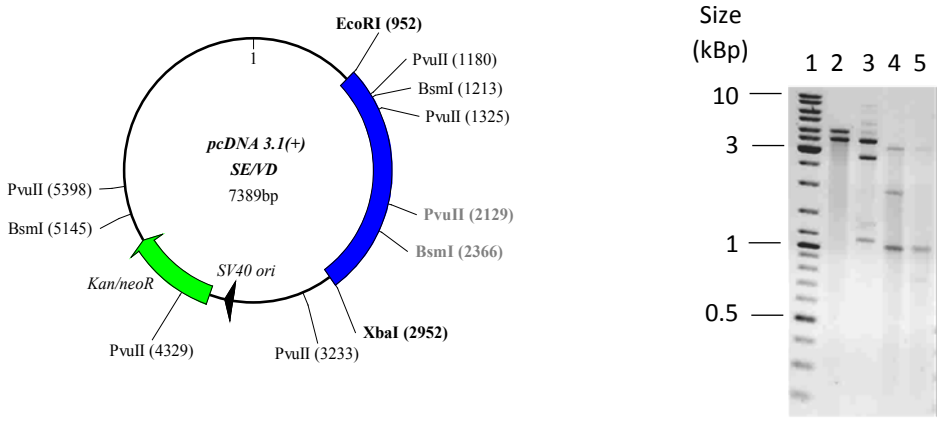


Figure 2.8: Restriction enzyme digests of the pcDNA 3.1+ SE/VD mutant.
Screening enzyme sites introduced by mutagenesis are indicated in grey on the vector map. Electrophoresis gel illustrates the difference in banding pattern between mutant and wild-type constructs. Lane 1: O'GeneRuler™ DNA Ladder Mix, lanes 2 & 4: pcDNA 3.1+ N-domain, lanes 3 & 5: pcDNA 3.1+ SE/VD. Lanes 2&3: *Bsm*I digests, Lanes 4&5: *Pvu*II digests.

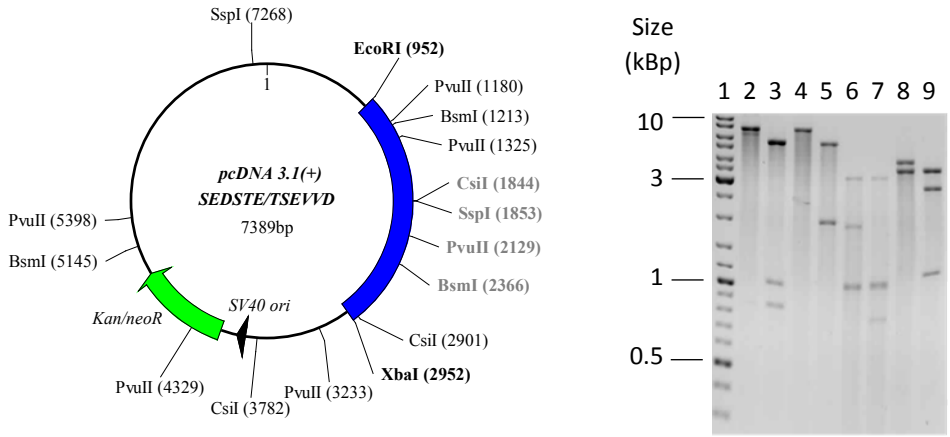


Figure 2.9: Restriction enzyme digests of the pcDNA 3.1+ SEDSTE/TSEVVD mutant.
Screening enzyme sites introduced by mutagenesis are indicated in grey on the vector map. Electrophoresis gel illustrates the difference in banding pattern between mutant and wild-type constructs. Lane 1: O'GeneRuler™ DNA Ladder Mix, lanes 2, 4, 6, 8: pcDNA 3.1+ N-domain, lanes 3, 5, 7, 9: pcDNA 3.1+ SEDSTE/TSEVVD. Lanes 2&3: *Csi*I digests, Lanes 4&5: *Ssp*I digests, Lanes 6&7: *Pvu*II digests, Lanes 8&9: *Bsm*I digests.

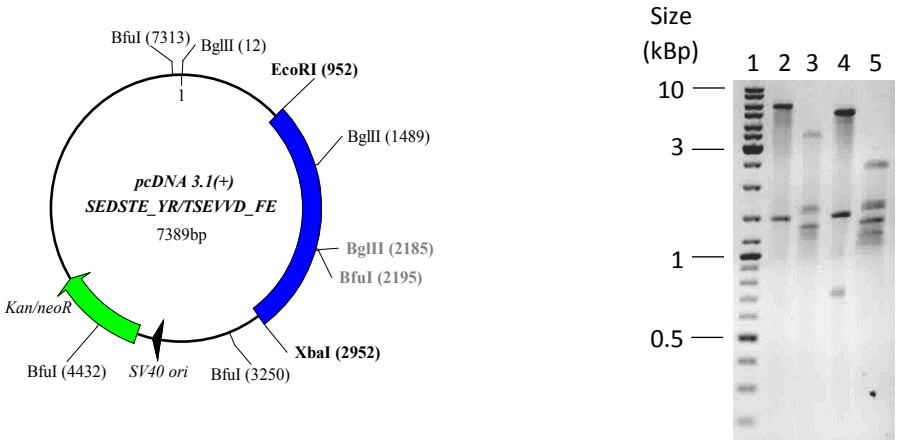


Figure 2.10: Restriction enzyme digests of the pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE mutant.
Screening enzyme sites introduced by mutagenesis are indicated in grey on the vector map. Electrophoresis gel illustrates the difference in banding pattern between mutant and wild-type constructs. Lane 1: O'GeneRuler™ DNA Ladder Mix (Fermentas), lanes 2-3: pcDNA 3.1+ N-domain, lanes 4-5: pcDNA 3.1+ SEDSTE_YR/ TSEVVD_FE. Lanes 2 and 4: *Bgl*II digests, Lanes 3 and 5: *Bfu*I digests.

2.6.2 Generation of a minimally glycosylated SEDSTE_YR/TSEVVD_FE construct

To enable crystallization of the mutant having all unique N-domain active site residues converted to C-domain, glycosylation sites n1, n2, n4, n5, n6 and n7 were removed. A *Bsu*36I/*Eco*RI digest was used to ligate the C-terminal part of the pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE construct (containing the active site mutations) to the N-terminal part of pcDNA 3.1+ N-domain n3 (having all glycosylation sites except n3 removed) (Figure 2.11).

Since site 7 is located between the active site mutants E262S and S357V and no restriction sites were available in this region for selective removal of site n7, site-directed mutagenesis was employed. PCR with the N289Q primer simultaneously introduced an *Nde*I site. The difference in banding pattern observed upon *Nde*I digest of the resulting amplicon and template DNA indicated positive incorporation of the desired mutation (Figure 2.12).

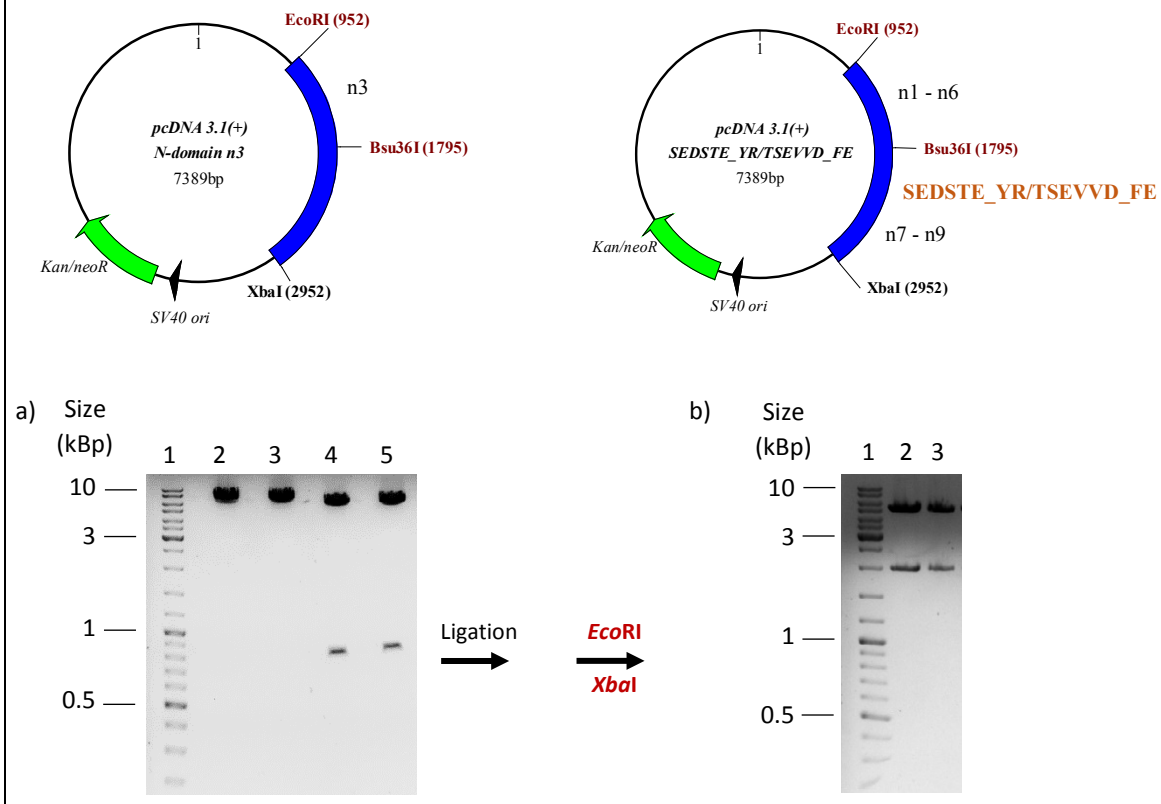


Figure 2.11: Sub-cloning used to construct pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n3789.
a) Lane 1: O'GeneRuler™ DNA Ladder Mix, lanes 2 and 5: pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE, lanes 3 and 4: pcDNA 3.1+ N-domain n3. Lanes 2-3: *Bsu*36I digests, Lanes 4-5: *Eco*RI/*Bsu*36I digests. b) Coding region size confirmation following ligation. Lane 1: O'GeneRuler™ DNA Ladder Mix, lane 2: pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE (fully glycosylated), lane 3: pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n3789, lanes 2-3: *Eco*RI/*Xba*I digests.

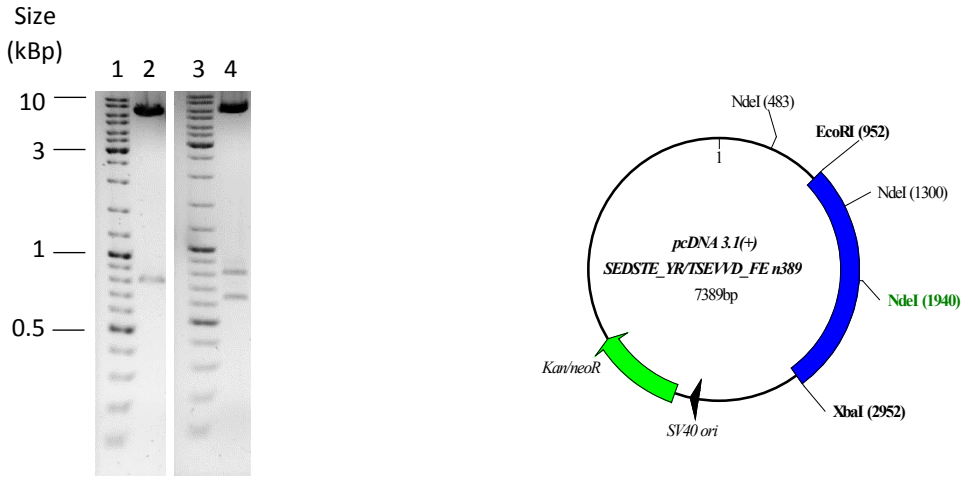


Figure 2.12: Restriction enzyme screen of pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n389 for positive removal of glycosylation site 7.
Lanes 1 and 3: O'GeneRuler™ DNA Ladder Mix, lane 2: pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n3789 (PCR template), lane 4: pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n389 (PCR amplicon). Lanes 2 and 4: *Nde*I digests.

Sequencing showed the presence of all SEDSTE_YR/TSEVVD_FE active site mutations, glycosylation sites 3, 8 and 9 and the absence of glycosylation sites n1, n2, n4, n5, n6 and n7. It, however, also led to the discovery of a spurious W40L mutation which resulted in the loss of a Pvu*II* restriction site. Further investigation through Pvu*II* digests of all constructs used in creating pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n389 led to the discovery of the spurious W40L mutation in the original pcDNA 3.1+ N-domain n3 template (data not shown). Since the N-domain n389 construct routinely used for crystallization does not contain this mutation, pBS SKII+ N-domain n389 was used to replace the N-terminus of pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n389 (with W40L) using sequential *Bsu*36I and *Eco*RI digests, ligation and *Eco*RI/*Xba*I screens as previously described (data not shown). Sequencing of the final construct showed the presence of W40, all eight active site mutations, glycan sites n3, n8 and n9 and N>Q mutations at the remaining glycan sites.

2.6.3 Purification of the active site mutant for X-ray crystallography

The pcDNA 3.1+ SEDSTE_YR/TSEVVD_FE n389 construct was stably transfected into CHO-K1 cells, heterologously expressed and the protein purified via lisinopril-sepharose affinity chromatography. Subsequent SDS-PAGE analysis showed a single band of ~70kDa for the N-domain n389 (Anthony *et al.*, 2010) and SEDSTE_YR n389 (Figure 2.13), indicating protein purity and a lower degree of glycosylation due to removal of sites 1, 2, 4, 5, 6 and 7. The protein was concentrated to 3.8mg/ml, frozen at -20°C and sent to Prof K. R. Acharya at the University of Bath for crystallization in an apo form as well as complexed to 33RE. An interpretation of the X-ray crystallography results will be given in section 2.6.12.

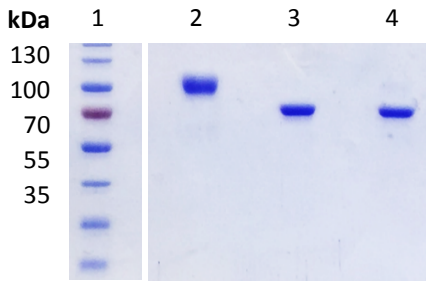


Figure 2.13: Coomassie-stained SDS-PAGE gel (10%) of the purified minimally glycosylated proteins. Lane 1: protein ladder (kDa), lanes 2, 3 and 4: fully glycosylated N-domain, minimally glycosylated N-domain n389 (Anthony *et al.*, 2010) and SEDSTE_YR n389.

2.6.4 Purification and kinetic characterization of the N-domain active site mutants

The N-domain, C-domain, SE/VD, SEDT/TSEV, SEDSTE/TSEVVD, SEDSTE_YR/TSEVVD_FE and YR/FE mutant constructs (referred to by the mutated N-domain residues from this point onwards) were stably transfected into CHO-K1 cells, and proteins purified to homogeneity via lisinopril-coupled sepharose affinity chromatography following heterologous expression. Single bands were observed for each construct on SDS-PAGE corresponding to approximately 90kDa for N-domain/mutants and 70kDa for C-domain (Figure 2.14).

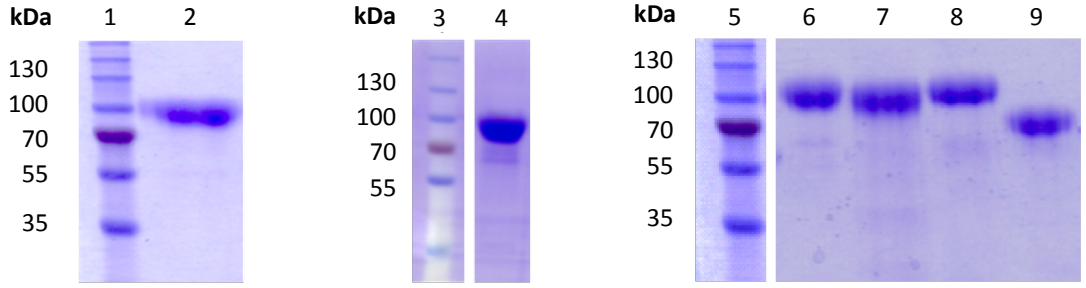


Figure 2.14: Coomassie-stained SDS-PAGE gels (10%) of purified fully glycosylated proteins.
Lanes 1, 3 and 5: protein ladder (kDa); lanes 2, 4, 6, 7, 8 and 9: WT N-domain, SE, SEDT, SEDSTE, SEDSTE_YR and WT C-domain.

The N-domain mutations introduced form part of the active site and as such could affect not only the binding of inhibitor but also substrate. The inhibitor (33RE) competes with substrate for binding to the active site, thus it is crucial to determine the effect of the mutations on the substrate's binding affinity as this could affect the observed inhibitor binding affinity.

Although catalytic turnover (k_{cat}) was altered upon mutation, the substrate Abz-FRK(Dnp)P-OH binding affinity (K_m) was comparable between the mutant and wild-type N- and C-domains (Table 2.3). Residual hydrolysis of this substrate could thus be used to accurately determine all the inhibitor binding constants. The overall catalytic efficiency (k_{cat}/K_m) was slightly higher for the wild-type C-domain compared to N-domain, consistent with previous reports (Araujo *et al.*, 2000).

Table 2.3: Catalytic efficiency of Abz-FRK(Dnp)P-OH hydrolysis by wild-type and mutant proteins.
Data was analysed using the Michaelis-Menten model for determination of K_m and k_{cat} values (GraphPad Prism® v6.01),

Construct	K_m (µM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}. μM^{-1})
N-domain	6.09	0.299	0.049
SE	10.76	0.010	0.001
SEDT	3.55	1.513	0.426
SEDSTE	7.33	1.050	0.143
SEDSTE_YR	5.76	3.770	0.654
C-domain	4.87	0.751	0.154

2.6.5 Kinetic characterization of ACE inhibition by 33RE

The affinity of 33RE and RXP407 binding to wild-type and N-domain mutant proteins were determined by means of kinetic assays using the non-domain-selective fluorogenic (Abz)-FRK-(Dnp)-P substrate. Dixon plots were constructed for wild-type as well as mutant enzymes to study the effect of active site mutations on inhibitor binding and thus selectivity. A representative plot is given in the Appendix A12.

Compound 33RE displayed low nanomolar inhibition of the N-domain (K_i=11.21 ± 0.74nM) and micromolar inhibition of the C-domain (K_i=11 278 ± 410nM), thus three orders of magnitude N-selectivity, similar to RXP407 (Table 2.4). To better represent the influence of mutation on 33RE selectivity, K_i values were compared on a logarithmic scale relative to the wild-type (Figure 2.15). Higher values indicate a decrease in inhibitor binding affinity upon mutation.

The N-domain S_2 mutants Y369F and R381E showed a drastic decrease in 33RE affinity (K_i=404.40 ± 17.30nM and K_i=86.97 ± 6.29 nM, respectively) with substitution of both these residues in the YR/FE mutant having an additive effect (K_i=2 794 ± 156nM). Together with the weak side-chain density observed for R381 (PDB ID: 4BXK) this suggests that alternative R381 locations are possible. The effect of these mutations on 33RE binding was similar to what was previously observed for RXP407 with the exception of YR/FE having a synergistic effect for RXP407 (Kroger *et al.*, 2009). Even though these S_2 residues do not account for the full selectivity, they clearly play a major role (Figure 2.15).

The S_2' mutant SEDT displayed a modest 2-3-fold decrease in N-selectivity (K_i=29.21 ± 2.40nM) while the S_2' double mutant SE (K_i=481.2nM) and the complete S_2' mutant SEDSTE (K_i=237.22 ± 13.60nM) dramatically decreased the selectivity by 43- and 21-fold, respectively (Table 2.4). Since these residues do not directly interact with 33RE, the prime mutants' kinetic results could not be rationalized from the N-domain-33RE crystal structure. Interestingly, the S_2/S_2' double pocket mutant SEDSTE_YR, having all the aforementioned mutations combined, abolished all selectivity (Table 2.4). The N-selective nature of ACE inhibition by 33RE thus appears to be governed by synergistic interactions with both the S_2 and S_2' subsites since lack of either resulted in decreased selectivity. It is inferred that the P_1 and P_1' groups have no contribution towards selectivity though they might be required for potency (Dive *et al.*, 1999).

Table 2.4: Affinity of 33RE binding to the wild-type and mutant proteins.
Inhibitor binding constants (K_i values) were determined from inhibition assays via Dixon plots and are represented as mean ± SD of two independent experiments, performed in triplicate.

	RXP407		33RE	
Construct	**K_i (nM)**	**Fold selectivity**	**K_i (nM)**	**Fold selectivity**
N-domain	21.03 ± 0.27	2 896	11.21 ± 0.735	1 006
C-domain	60 826 ± 9 175		11 278 ± 410	
Y369F			404.4 ± 17.3	27.9
R381E			86.97 ± 6.29	129.7
YR			2 794 ± 156	4.04
SE		*n.d.*	481.2 †	23.4 †
SEDT			29.21 ± 2.40	386.1
SEDSTE			237.22 ± 13.60	47.5
SEDSTE_YR			10 009 ± 157	1.13

n.d.: not determined in the current study (see Kröger *et al*, 2009 (Kroger *et al.*, 2009) for S_2 mutants' RXP407 kinetics); † One experiment in triplicate due to poor protein expression and decreased catalytic efficiency

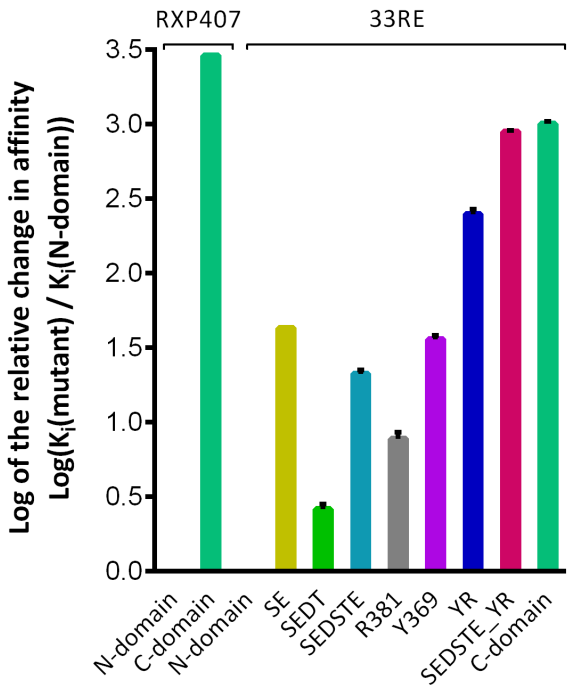


Figure 2.15: Logarithmic scale comparison of the relative inhibitor binding affinity of N-domain mutants to that of the wild-type domains.
Relative binding affinities for RXP407 and 33RE are shown on the left and right, respectively. Positive values represent decreased affinity relative to wild-type N-domain, thus towards more C-domain-like K_i. Results are shown as mean ± SD from two independent assays performed in triplicate.

2.6.6 Influence of mutation on 33RE tetrazole stability and interaction energy

Structures of the various mutants prepared and kinetically analysed *in vitro* were also created *in silico* to simulate their dynamic behaviour in the presence and absence of ligand. Prior to use of the wild-type N-domain crystal structure complexed to 33RE (PDB ID: 4BXK) in simulations, it was re-refined against its electron density. This led to a change in initial and final values for R_{free} of 0.1814 and 0.1784, respectively; RMS(bonds) of 0.020 and 0.006, respectively; and RMS(angles) of 1.322 and 0.916, respectively. The resulting structure was prepared as described in the methods section for the wild-type N-domain, SE, SEDT, SEDSTE, SEDSTE_YR and YR mutants in apo, 33RE-bound and 33RE P_2' carboxylate analogue-bound forms.

The wild-type and mutant structures were explicitly solvated and neutralized by adding approximately 17500 water molecules, 53 sodium and 47 chloride ions. Each structure thus consisted of approximately 62700 atoms. Two thousand five hundred and ten trajectory frames were collected for all apo and ligand-bound proteins from 12550ps of production dynamics. The first 3000ps was discarded as part of equilibration since the structures relaxed and showed

further decreases in potential energy over this timeframe. Trajectories were analysed for the remaining 9550ps, sampled at 2 frame intervals for a total of 955 conformations per structure.

The residues' root mean square deviation (RMSD) was calculated across these 955 frames with backbone fitting to the first frame (post-equilibration) to remove the rotational and translational motion. All the structures were stable during this timeframe (Figure 2.16) and could be analysed further.

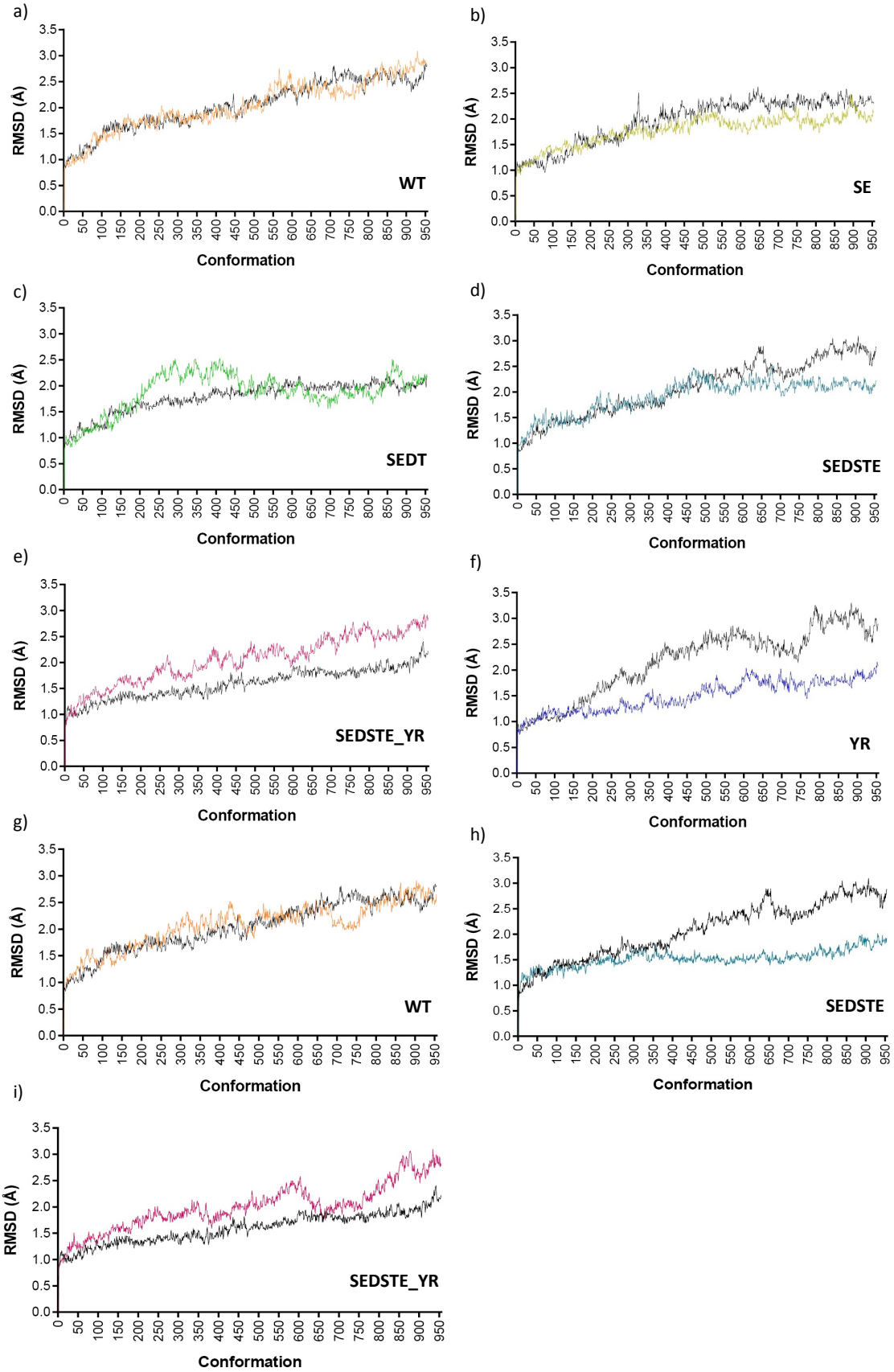


Figure 2.16: Root mean square deviation (RMSD) over the 9550ps of simulation (last 955 conformations) after backbone fitting against the first equilibrated frame.

Apo structures are shown in black with their corresponding 33RE-bound (a-f) or analogue-bound (g-i) structures in colour a) and g) WT, b) SE, c) SEDT, d) and h) SEDSTE, e) and i) SEDSTE_YR and f) YR.

The secondary structure nomenclature to be used from this point onwards is according to the 33RE co-crystal structure (PDB ID: 4BXK) assignments from the European Bioinformatics Institute PDBSum database records (Figure 2.17).

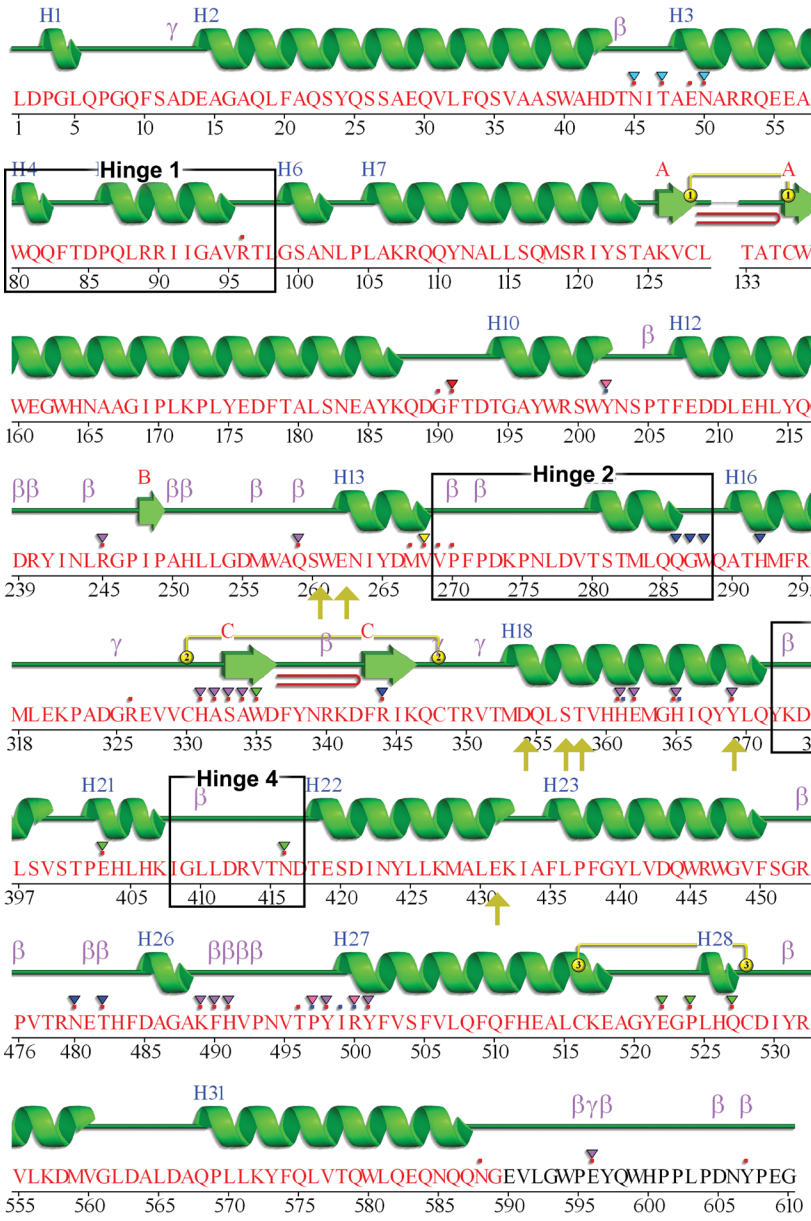


Figure 2.17: Secondary structure of the N-domain (PDB ID: 4BXK).
Beta sheets are labelled as A, B or C, helices are labelled by number, the proposed hinges are indicated by boxes and sites of the mutations studied by yellow arrows.

In wild-type, binding of 33RE's amide was seen to occur with the conserved K489, H491 and Y498 residues, similar to the interaction of most ACE inhibitors' carboxylate in this subsite. Although the crystal structure showed binding to the conserved Q259 as well, the simulations only showed Q259 transiently binding to K489 and possibly indirectly assisting in its binding to 33RE's amide.

Consistent with the 33RE co-crystal structure, the MD simulations failed to show any direct interactions to the unique N-domain S_2' residues or R381 in the S_2 subsite. Per-residue decomposition of interaction energies calculated for all frames and averaged over the 10ns simulation time showed that the majority of the binding affinity resulted from pi-stacking and Van der Waals interactions to H331, A332, S333, W335 and H491 (Figure 2.18). Some contributions to binding affinity also arose from hydrogen bonding of the 33RE backbone to A334, the amino group to the unique S_2 residue Y369, the phosphonate group to the conserved Y501 and the amide to the conserved K489 residue. The positive interaction energy to the zinc-binding E389 is likely as a result of electrostatic repulsion by the phosphonate. Since the coordination of zinc to a phosphonate, two histidines and a glutamate has not been parameterized before, distance restraints were applied to maintain zinc-coordination throughout the simulation. Without the use of custom partial charge values derived from quantum mechanical calculations the negative charge of the phosphonate was likely not shared by the positively charged zinc ion, resulting a higher negative charge and repulsion of E389. Although this precludes calculation of the net binding energy of the inhibitor, interaction energies between moieties other than the phosphonate could be calculated. Taken together, the P_2' amide's hydrogen bonding to conserved loop residues of subdomain 2 and the aminomethyl tetrazole binding to Y369 and stacking to beta sheets C of subdomain 1 synergistically enabled the bridging of 33RE across the active site cleft (Figure 2.19).

Mutation only slightly affected the energies of these interactions with SEDSTE_YR showing the weakest binding due to a 43-94% decrease in interaction energy (versus wild-type) to Y369, F435, K489, H491, T496, Y498, R500 and Y501 (Figure 2.18 b). Considering the ten average ligand poses calculated over 1ns intervals each, however, it was found that while the tetrazole is very stable in the WT N-domain structure, its stability is decreased and orientation changed upon mutation (Figure 2.18 c-h). This ultimately results in suboptimal bridging of 33RE across the mutant active site cleft (Figure 2.19 b).

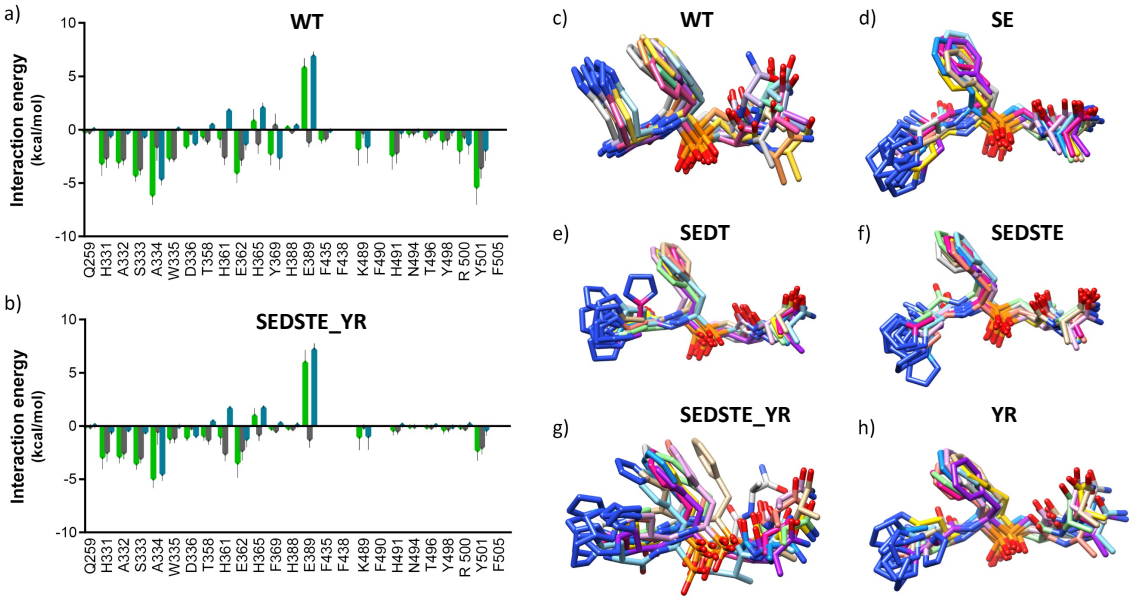


Figure 2.18: Interaction energies and ligand poses calculated over all frames of the 33RE-bound structures.
Total interaction energy (green), Van der Waals component (grey) and electrostatic component (blue) represented as mean ± standard deviation for a) WT and b) SEDSTE_YR. Ligand poses of 10 average superimposed structures over 1ns each for c) WT, d) SE, e) SEDT, f) SEDSTE, g) SEDSTE_YR and h) YR.

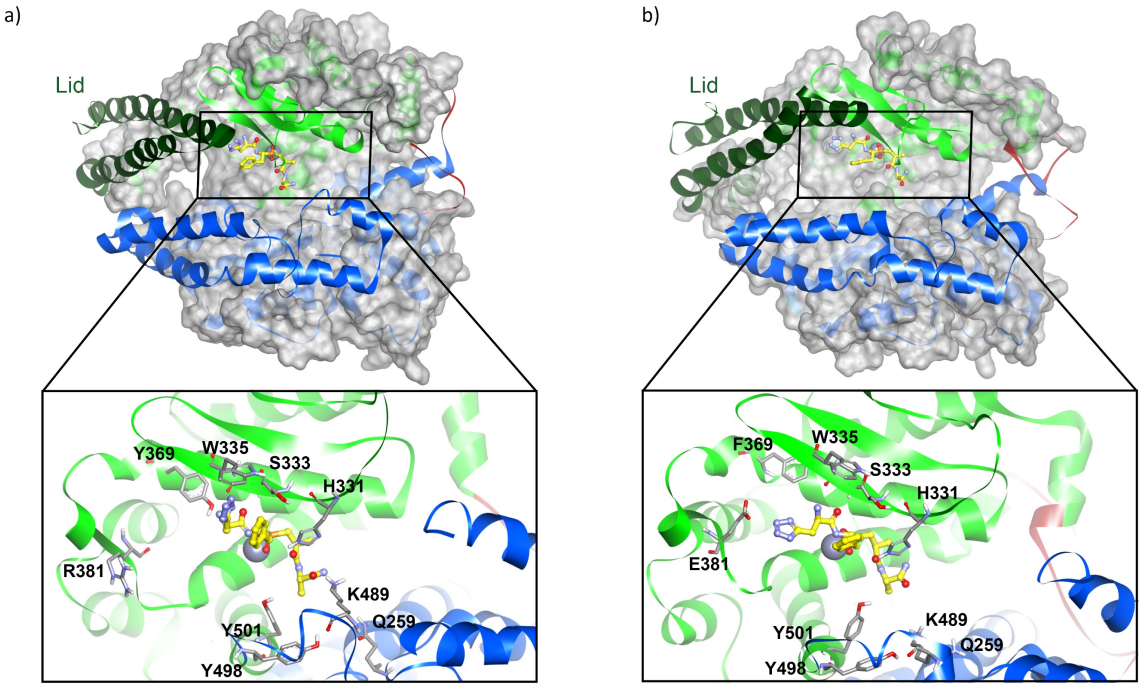


Figure 2.19: Bridging of 33RE across the active site cleft.
Subdomain 1, subdomain 2 and the lid flanking the cleft are shown in green, blue and dark green, respectively. Interacting residues are shown in grey sticks with 33RE in yellow for the average conformation from the last 1ns simulation of a) WT and b) SEDSTE_YR.

2.6.7 The effect of active site residue mutations on apo protein flexibility

Since alterations in ligand stability can result from changes in protein flexibility, the amino acids' root mean square fluctuation (RMSF) was calculated across the 955 frames and compared to wild-type. Both the apo and bound forms were studied to ascertain whether mutation affected the dynamic nature of the protein prior to inhibitor binding or the stability of the resulting complex.

Apo structure flexibility

Firstly, looking at the apo forms it is noted that upon mutation some changes in backbone flexibility occur around the lid helices 3-5 (hinge 1), beta sheets A, hinge 2, hinge 3 and hinge 4 (hinge nomenclature according to the proposed regions described from normal mode analysis by Watermeyer *et al* (Watermeyer *et al.*, 2006)) (Figure 2.20).

SE mutation decreased the flexibility of hinges 2 and 4 but increased hinge 3 flexibility over wild-type with no great changes in the lid (Figure 2.20 a). The hinge region 3 was unaffected in all other apo structures. Mutation of SEDT decreased the lid, hinge 2 and hinge 4 flexibility, though to a lesser extent than in SE (Figure 2.20 b). The combination of these in SEDSTE resulted in hinge 2 and 4 RMSF decreases similar to that of SE which could therefore be attributed mainly to these two residues (Figure 2.20 c). In addition, the flexibility is increased for helix 10, the 450-460 loop and the N-terminal, prime side of the lid's helix 3. Mutation of YR in the non-prime subsite did not decrease the hinge flexibility and showed along with increased RMSF of the N-terminal, prime side of helix 3, decreased RMSF of the C-terminal, non-prime side of helix 3 (Figure 2.20 e). Mutation of all the aforementioned residues in SEDSTE_YR led to a similar effect on the lid's helix 3 with additional decreased flexibility of hinges 2 and 4 (Figure 2.20 d). Even though the resulting S_2' mutant complexes' ligand interaction energies were comparable, changes in the apo protein dynamics upon mutation might thus have affected the initial affinity for inhibitor binding.

Directions of motion affecting N-domain distal hinge and lid regions

Further insight into the directions of motion and the mechanism whereby these active site mutations affect the distal hinge and lid regions was gained by using a combination of different MD analyses. Firstly, the most distinct conformations were identified from all-frame RMSD

clustering heatmaps, aligned and coloured according to RMSF (blue to red = low to high RMSF). Secondly, the relationship between the mutated residues and lid and hinge flexibility was studied by identifying correlated protein backbone motions from covariance heatmaps. Diagonalization of the covariance matrices by PCA allowed inspection of the long range, low frequency motions conserved throughout the simulation. In all structures the first five principal components accounted for 50-60% of the overall motion (scree plots not shown). The first principal component, which describes the majority of this, was compared across these complexes with the magnitude and direction of the eigenvalues visualized through porcupine plots.

From these analyses the wild-type apo structure's breathing appeared to be mediated by the tandem motion of hinges 2 and 4 located parallel to each-other and between the two subdomains (Figure 2.20 f). This is accompanied by alternate lid movement through lateral shift of the lid's helices 2-3 over the cleft and outward tipping of the opposite end's helices 4-6 by twisting at the Y76 beta turn. SE mutation greatly decreased the hinge 2 and 4 flexibility and increased the separation between the subdomains in the prime subsite. The isolated nature of the two C-domain residues introduced into the N-domain on helix 18 (subdomain 1) and helix 22 (subdomain 2) likely resulted in instability of the adjacent hinge 3 loop between the subdomains, and thus its increased RMSF. Mutation of all the prime subsite residues was found to affect cleft closure by not only decreasing the hinge 2 and 4 flexibility but also altering the direction and magnitude of lid motion. Instead of a cleft-closing lateral shift, the mutant's helix 3 N-terminus was seen to roll over helix 2 and increase active site exposure.

Although the YR mutant displayed hinging similar to wild-type, R381E and the resulting salt-bridge formation to R90 of helix 5 fixed this end of the lid in a more inward position and enhanced the lateral shift of the opposite end at the N-terminus of helix 3 for active site cleft closure.

Lastly, the combined SEDSTE_YR mutant displayed increased prime subsite subdomain separation, decreased hinging, fixing of the lid at helices 4-6, enhanced flexibility of the N-terminus of helix 3 in a rolling motion and thus ultimately a more open state (Figure 2.20 g). Altogether, this suggests that the unbound C-domain might be more rigid/exposed and have a greater energy barrier to overcome for effective active site closure than the N-domain.

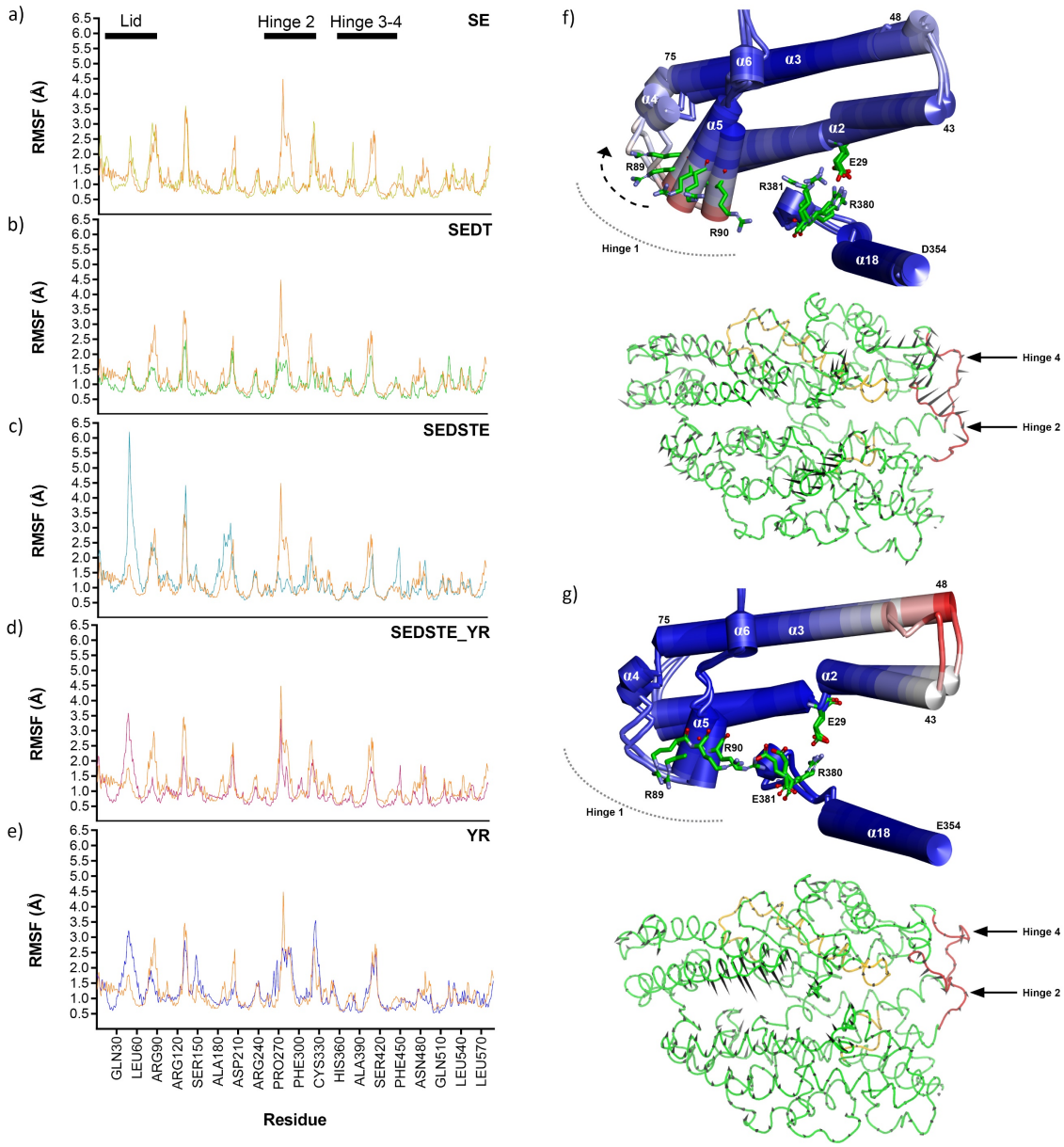


Figure 2.20: Apo structure flexibility.
Backbone atom root mean square fluctuation (RMSF) is given for a) SE (yellow), b) SEDT (green), c) SEDSTE (teal), d) SEDSTE_YR (magenta) and e) YR (navy) in relation to wild-type (orange). f) and g) Clusters of conformations coloured according to RMSF (low = blue, high = red, α = helix) for the lid region and porcupine plots of correlated C_α motions from the first principal component for WT and SEDSTE_YR, respectively. Quills indicate the direction and magnitude of motion. Hinges 2 and 4 are given in red and helices 13, 18 and 22 in yellow.

2.6.8 The effect of active site mutations on protein hinging upon 33RE binding

Wild-type 33RE-bound structure

The trend in backbone RMSF of the simulated wild-type 33RE-bound structure compared well to that calculated from experimental C_α B-factors of PDB ID: 4BXK using equation (1) (where B_i is the main chain B-factor of the i^{th} residue). The absolute values were, however, slightly lower as is to be expected for the crystal form of the protein at low temperature.

$$RMSF = \sqrt{\frac{3B_i}{8\pi^2}} \quad (1)$$

Comparing the wild-type apo and 33RE-bound forms, an increase in backbone lid flexibility was noted in the presence of 33RE (Figure 2.21). In the bound form the S_2' unique S357 formed a strong, stable hydrogen bond with the unique E431 that is maintained throughout the simulation. Residues S357 and E431 are located opposite each-other on helix 18 of subdomain 1 and helix 22 of subdomain 2, respectively, thereby forming a stable contact point at the interface of the two subdomains (Figure 2.21 f). The unique E262 (on loop between helices 12 and 13) was observed to hydrogen bond to the unique D354 backbone at the tip of helix 18 while the D354 sidechain formed a salt bridge to R350 on the adjacent loop. R350, however, also interacted with D140 on helix 8. This helix thus moved inwards towards R350 but was repelled by the negative charge from D354, E262 and E431, resulting in its high RMSF and opening/closing of the active site. The interaction of E262 with helix 18 and the presence of a favourable subdomain interface resulted in the movement of helices 12, 13 and the connecting loop towards the active site. The interaction network between E262, D354, R350, S357 and E431 (Figure 2.21 f) thus facilitated the correlated cleft-closing long-range motions of helix 18 (Figure 2.22 a) and fixed the loops of hinge 2 in more bent conformations (vs apo), likely contributing to the high RMSF and movement of helix 15 towards the active site opening.

In the non-prime subsite R381 located on hinge 3 slightly transitioned into the active site but did not bind the tetrazole. This was likely the result of R381 backbone (and hinge 3 loop) stabilization through hydrogen bonding to the H388 sidechain. Although a tetrazole can interact with arginine it would be sterically hindered in the S_2 and likely weaker due to the tetrazole size and less optimal bond angle (Tominey *et al.*, 2006) (compared to the Asp carboxylic acid in RXP407 where R381 is positioned in the active site). Movement of the R381 sidechain facing

out of the active site caused charge repulsion of the conserved R90 residue of the lid's helix 5 (Figure 2.21 f). This resulted in an upward tipping motion of the C-terminus of helix 3 and helices 4-6 about a pivot point formed by an R380-E29 salt bridge (E29 is located in the middle of the lid on helix 2). This, in turn, resulted in the downward tipping of the opposite end of the lid at the N-terminus of helix 3 towards the prime subsite. This see-saw motion of the lid in the bound as opposed to lateral shift of the unbound further enhanced the active site cleft closure. Covariance analysis showed a clear positive correlation between residues of helix 15-20 and the N-terminus of helix 3 (subdomain 1) (Figure 2.22 a). These residues are negatively correlated to helices 4-6 (lid R90), helix 8 (D140, subdomain 2) and helices 12-13 (E262, subdomain 2). The unique S_2 and S_2' active site residues of the N-domain thus acted synergistically to bring about a concerted active site closing motion through hinges 1, 2, 3 and 4.

Mutant 33RE-bound structure

In contrast to the situation of the apo forms, in the presence of 33RE the flexibility of hinges 2 and 4 were not affected by mutation. There was, however, a distinct decrease in flexibility of the helices 12-13 and their connecting loop (Figure 2.21) since none of the prime mutants possessed the E262 interaction with helix 18 to facilitate the inward motion of this region. The negative correlation with helix 18 was thus also lost.

SE mutation

The SE mutation decreased the flexibility of the lid at helices 3-5, beta sheets A and helix 8 but increased the flexibility of hinge 3. Upon mutation, the hydrophobic repulsion between S357V and E431D increased the distance between helices 18 and 22, thereby decreasing the repulsion of D140 (helix 8) by D354 and E431D. The lack of S357 binding and shorter chain length of D431 allowed direct hydrogen bonding of D431 to E262. Binding of the D354 backbone by E262 thus did not occur and likely contributed to the separation of helix 18 from subdomain 2. The D140-R350 and D354-R350 bonds are thus more stable with a decrease in helix 8 backbone RMSF (Figure 2.21 a). Yet, the lack of binding between the prime residues on helices 18 and 22 (S357-E431 in WT) and the isolated nature of these mutations decreased the helix 18 stability, thereby increasing the flexibility of Y369 and the hinge 3 loop harbouring R381. Since Y369 binds the aminomethyl tetrazole moiety in wild-type, this increased flexibility likely led to decreased tetrazole stability (Figure 2.18 d). RMSF of R380 was also increased which decreased the lid's

E29-R380 pivot point stability and along with the SE subdomain contact point mutation made the lid's see-sawing motion less likely so that the structure is overall more exposed and static.

SEDT mutation

The SEDT mutation increased the lid flexibility at the N-terminal prime side of helix 3 but decreased the flexibility of beta sheets A and helix 8 (Figure 2.21 b). This mutation still allowed lid see-sawing motion since the SE is still present to form an interface between the subdomains. The T358V mutation, however, appeared to repel the D354E residue sidechain since valine is hydrophobic and the polar glutamic acid longer. E354 therefore changed direction and bound to K427 on helix 22 opposite, thus forming a strong salt bridge to helix 18. E354 also hydrogen bonded to E431 and therefore the S357-E431 interaction was lost, but helix 18-22 subdomain contact maintained. This caused a slight shift in the position of helix 18 and the altered E354 orientation allowed formation of a stable R350-D140 bond. This, and the lack of repulsion by acidic prime residues, decreased the RMSF and long-range motions of helix 8. The lid motion at the N-terminus of helix 3 was increased over WT since the slight helix 18 shift created more space for the lid to swing in. The hinge 3 RMSF was unaffected by SEDT and therefore the E29-R380 pivot was stable and R381 repulsion of R90 facilitated lid tipping.

SEDSTE mutation

In the case of SEDSTE it was found, similar to SE, that flexibility was decreased at the lid helices 3-5, beta sheets A, helix 8 and the loops between helices 12 and 13 – no effect was seen on hinge 3, however (Figure 2.21 c). The mutation of T358V and S357V created a hydrophobic patch along with L356 and V359. This resulted in the D354E sidechain changing orientation towards K427 and formation of a stable salt bridge. Yet, the increased side-chain length of D354E and K427 precluded close interaction between helices 18 and 22. The decreased side-chain length of E431D also prevented the E354-D431 binding as seen for SEDT. The lack of D354E-R350 interaction and decreased negative charge of the prime subsite increased the D140-R350 interaction's stability with a resulting decrease in helix 8 flexibility and long-range motion towards the active site opening. The six prime mutations together therefore exposed the active site to the exterior environment by increasing the subdomain separation between helices 18 and 22 in this subsite.

YR mutation

In the nonprime subsite YR mutation decreased the flexibility of hinge 2, the loop between helices 12 and 13, the active site helix 18, the lid's C-terminal non-prime side of helix 3, and the region spanning helices 4-9 (Figure 2.21 e). Important to note is the decreased flexibility of the 60-95 region, but unaffected 1-60 lid region, which is only found for the YR and SEDSTE_YR 33RE-bound forms. Mutation of Y369F with hydrophobic repulsion of the tetrazole led to its movement towards H388 and disruption of the H388 hydrogen bonding with the hinge 3 backbone. In YR the R381E thus electrostatically interacted with the adjacent R380. Repulsion of R90 thus did not occur and it instead bound the R381E backbone, decreasing the helix 4-6 RMSF. This stabilization of the non-prime subsite along with the favourable prime subsite interface enhanced the active site closure.

SEDSTE_YR mutation

The R90 repulsion observed in the YR mutation did not occur in SEDSTE_YR where R381E was located in the active site (Figure 2.21 g). The added prime subsite mutations in the presence of 33RE, however, increased the hinge 3 flexibility likely synergistically through increased helix 18 - 22 separation, R381E electrostatic repulsion by the unstable, electronegative tetrazole and hydrophobic repulsion by Y369F. Although the lid's helix 3 displayed wild-type RMSF (Figure 2.21 d) and positive correlation to helix 18 (Figure 2.22 b), the direction of motion was changed so that it rolled outwards over helix 2 instead of tipping towards the active site cleft as in wild-type and YR. Altogether, the increased helix 18 - 22 separation, lack of helix 13 E262-D354 (C_a) binding and the altered lid motion resulted in a more exposed active site cleft.

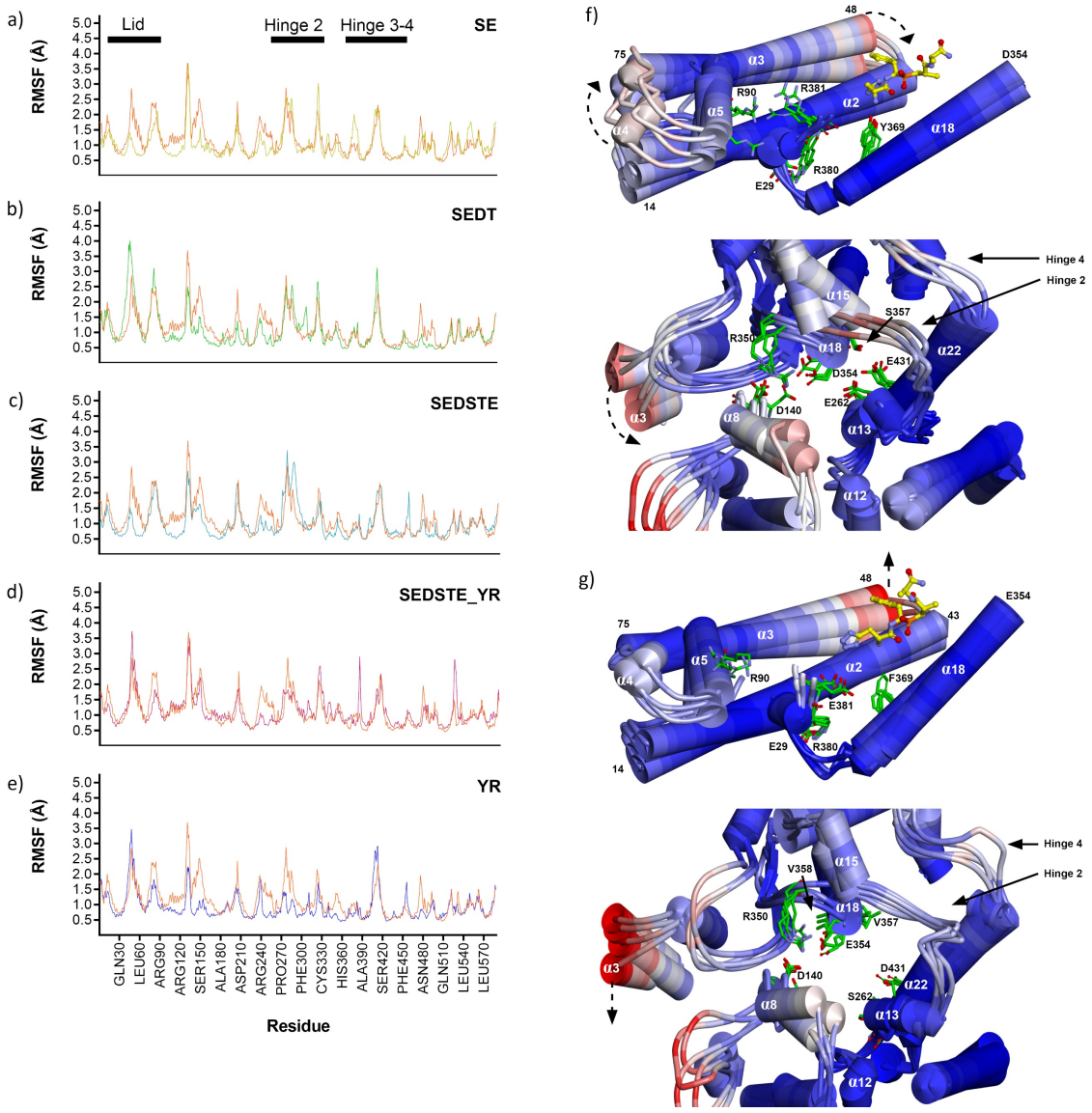


Figure 2.21: Backbone atom root mean square fluctuation (RMSF) of the 33RE-bound structures.
a) SE (yellow), b) SEDT (green), c) SEDSTE (teal), d) SEDSTE_YR (magenta) and e) YR (navy) in relation to wild-type (orange). The lid and prime subsite interactions for the clustered WT and SEDSTE_YR are shown in f) and g), respectively with low to high RMSF coloured from blue to red, protein carbon atoms in green and 33RE in yellow (α = helix).

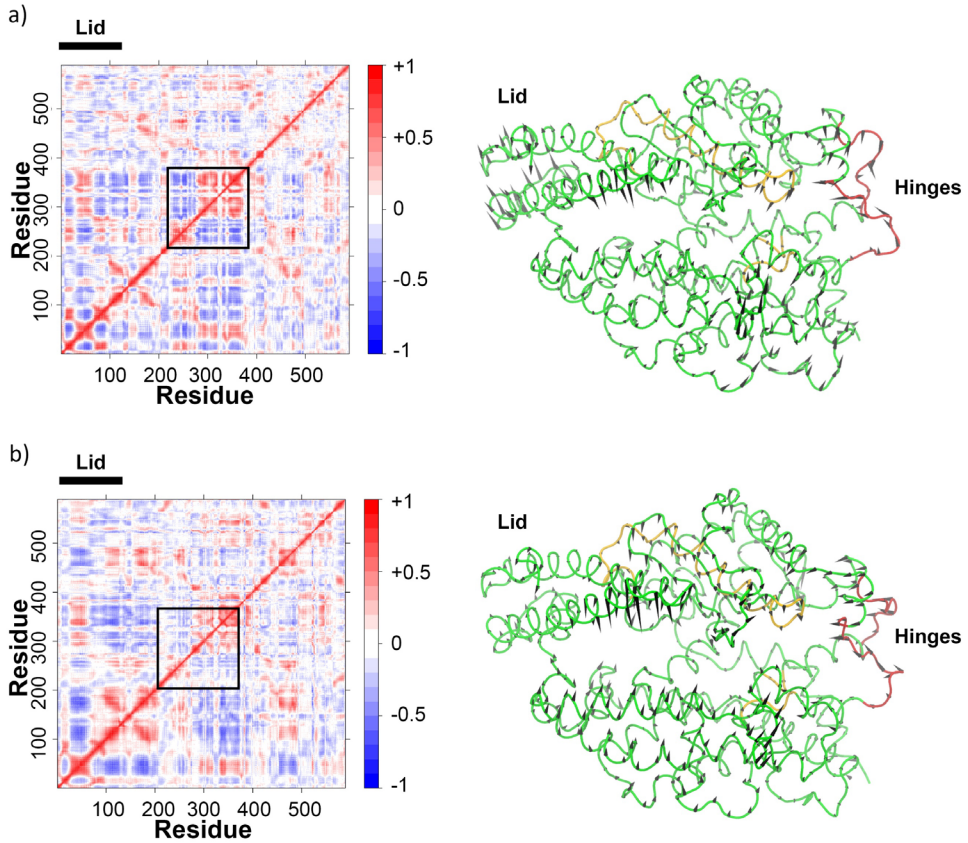


Figure 2.22: Correlated motions in 33RE-bound structures.
Heatmaps indicate positive (red), negative (blue) and no correlation (white) between C_α motions. Directions of these motions are shown on the corresponding porcupine plots of the first principal component for a) WT and b) SEDSTE_YR. Helices 13, 18 and 22 are given in yellow, hinges 2 and 4 in red and the approximate location by the boxes.

2.6.9 Assessment of active site contraction upon 33RE-mediated protein hinging

To ascertain whether the mutations' effect on hinging and lid motion indirectly affected ligand binding affinity by influencing the transition from open, apo forms to more closed, ligand-bound forms, the average structure of each apo was aligned to its corresponding average bound form. As an indication of how the overall protein compactness varied over time, the radius of gyration and the central active site cavity volume was determined for 10 average structures calculated over 1ns intervals for each 10ns MD simulation.

The wild-type N-domain was highly dynamic prior to ligand binding with its central cavity volume ranging between 11500 $Å^3$ and 18500 $Å^3$ (Figure 2.23a). Although the protein was still dynamic upon binding of 33RE, its radius and cavity volume was decreased with the poorly aligned active

site helices 18, 13 and 22 suggesting aid of their movements in ligand binding, subdomain interaction, cleft closure and stabilization of the complex.

Mutation of the S_2' residues located at the subdomain interface abrogated this stabilization of the closed conformation with the isolated prime mutation SE causing cavity expansion upon 33RE binding and a less dynamic active site (narrower volume range) (Figure 2.23 b). The SEDT mutant, though also less dynamic, still displayed decreased volume and radius upon binding (Figure 2.23 c) while the SEDSTE, similar to SE, did not contract and showed increased subdomain separation upon binding. The active site contraction was thus likely a result of helix 18 motion towards helix 22, inter-subdomain contact through S357 and E431, and lid tipping. Since YR still possessed N-domain prime residues and the R381E mutation stabilized helices 4-6 towards the active site, its radius and volume was greatly decreased upon binding, despite 33RE's repulsion by F369 (Figure 2.23 f). Additional mutation of the prime residues, however, abrogated this contraction and the radius as well as volume was notably increased upon 33RE binding (Figure 2.23 e).

A similar active site closing motion was previously proposed for the ACE2 homolog where the open and closed crystal structure forms were poorly aligned in this prime subsite region (Towler *et al.*, 2004). Lack of subdomain interaction and thus increased active site exposure upon mutation has implications for ligand binding affinity since it also increased the 33RE solvent exposure (on average).

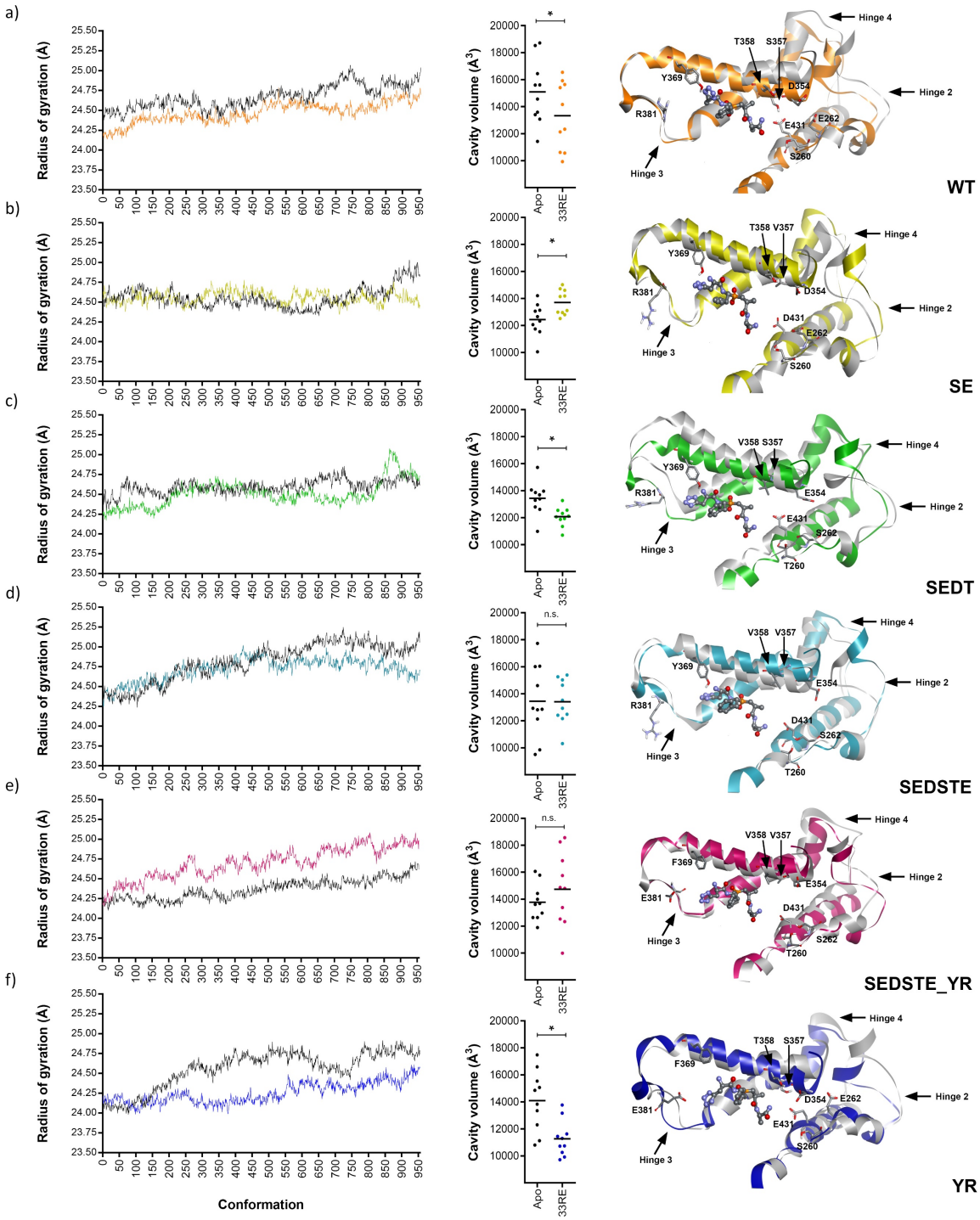


Figure 2.23: Volume analysis of 33RE-bound structures.
Radius of gyration over the last 955 frames, central cavity volume calculated over 1 ns sections (each represented by a dot with the mean as a line) and alignments of each overall average bound structure to its respective apo form are shown for a) WT (orange), b) SE (yellow), c) SEDT (green), d) SEDSTE (teal), e) SEDSTE_YR (magenta) and f) YR (navy) 33RE-bound structures in relation to their respective apo forms (black line/grey structure). Unique residues from the bound average structures are shown as sticks. Volumes of bound vs apo were compared with statistical significance set at a 99% confidence level (*P < 0.01; n.s.: not significant).

2.6.10 Differences in hinging between amide and carboxylate P_2' analogues

Looking at the 33RE P_2' carboxylate analogue it was noted that the wild-type structure was actually expanded more than in the apo structure upon binding (radius of gyration not shown) and no decrease in active site volume was detected (Figure 2.24). The interaction energy to K489 and H491 was thus similar to that of the amide with the expansion of the prime subsite through a shift in helix 18 likely leading to contraction of the non-prime at the adjoining hinge 3 and strengthening of the H331 and H388 interactions (Figure 2.24 a).

Upon mutation of SEDSTE the RMSF was decreased overall (data not shown), the interaction energy to K489 and H491 increased, the active site less exposed (despite repulsion between prime subsite residues) and the protein more compact (Figure 2.24 b). This suggests that the negatively charged inhibitor P_2' carboxylate is less favoured for binding to the electronegative N-domain prime subsite than the more neutral C-domain counterpart. The non-prime subsite was therefore not contracted and H331 binding was similar to 33RE. Since the P_2' carboxylate interaction is strong and the prime subsite effectively closed, the additional mutation of YR with resulting Y369F-tetrazole repulsion did not affect inhibitor stability as for 33RE. In this case the R90 binding by R381E and thus increased stability of helices 4-6 also increased the interaction energy to H331.

Selective active site closure and inhibition of the N-domain over the C-domain thus not only depends on the unique prime and non-prime residues but also the inhibitor's amide. This supports the hypothesis that the C-domain requires a stronger P_2' moiety, such as a carboxylate, for active site closure to occur by binding the loop between helices 12 and 13 and overcoming the energy barrier caused by repulsion at the subdomain interface. Although these simulations were done on 33RE which has the tetrazole in P_2, the binding affinity trend is similar to that presented for the RXP407 P_2' carboxylate analogue *in vitro* by Dive *et al* (Dive *et al.*, 1999) in that the P_2' carboxylate increased the binding affinity to both domains, thereby abolishing selectivity.

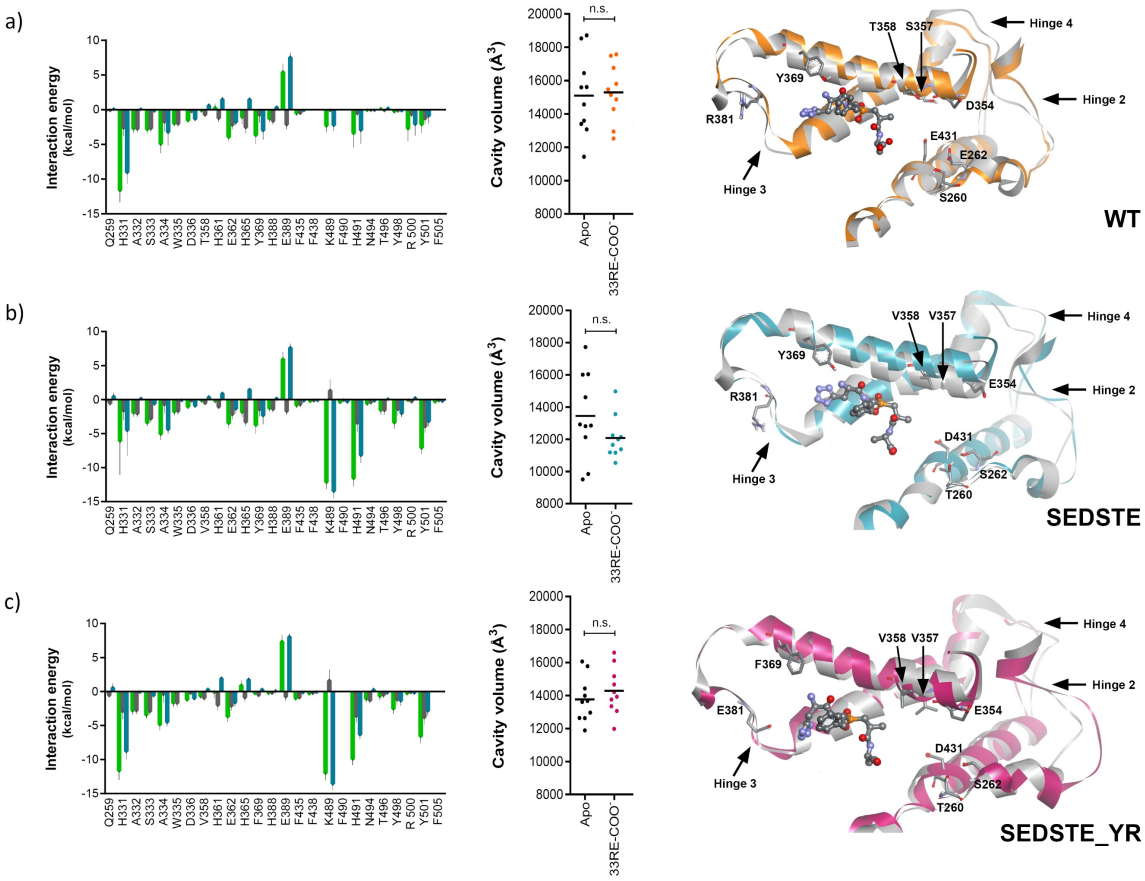


Figure 2.24: Interaction energies and volume analysis of the 33RE P_2' carboxylate analogue-bound structures.
Total interaction energies (green), Van der Waals components (grey) and electrostatic components (blue) were calculated over all frames and represented as mean ± standard deviation for a) WT, b) SEDSTE, and c) SEDSTE_YR. Central cavity volume calculated over 1 ns sections (each represented by a dot) and alignments of each overall average bound structure to its respective apo form are shown for WT (orange), SEDSTE (teal) and SEDSTE_YR (magenta) 33RE analogue-bound structures in relation to their respective apo forms (grey structure). Unique residues and the 33RE analogue pose from the bound average structures are shown as sticks.

2.6.11 *In vitro* assessment of change in active site solvent exposure upon 33RE binding

To evaluate the effect of mutation on protein stability and 33RE-mediated active site closure *in vitro*, differential scanning fluorimetry was performed and the protein melting temperatures (T_m) determined across a range of 33RE concentrations. This technique utilizes a dye binding to the hydrophobic protein core to allow detection of fluorescence signal upon protein unfolding with increasing temperature. The first derivative of the resulting signal was used to construct a melting curve with the T_m derived from the midpoint of the peak as described by Niesen *et al* (Niesen *et al.*, 2007) (Figure 2.25 a).

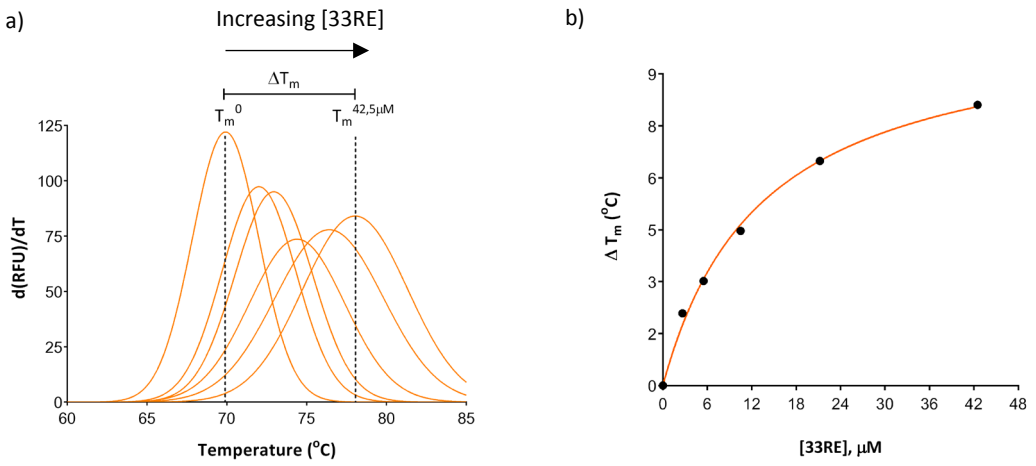


Figure 2.25: Differential scanning fluorimetry results of wild-type N-domain titrated with 33RE.
a) Melting curve from one representative assay showing an average gaussian fit over three wells at each 33RE concentration, obtained from the first derivative of the fluorescence signal; b) kinetics of thermal stabilization upon 33RE titration of 1 µM protein in triplicate wells. T_m^0 indicates protein melting temperature in the absence of 33RE; ΔT_m values were calculated from the rightward peak shift from T_m^0 on the derivative curves. T_m^{max} was defined as the melting temperature at the 33RE concentration beyond which no further stabilization was observed (plateau on b)). RFU = relative fluorescence units.

The determined N-domain apo protein melting temperature (T_m^0) of 69.5±0.6°C (Figure 2.26) compared well with the circular dichroism values of 70.5°C (Voronov *et al.*, 2002a) and 72°C (O'Neill *et al.*, 2008) for bovine and human N-domain, respectively, thus illustrating the reliability of this method in studying ACE thermal denaturation. Binding of a ligand to the protein active site is generally thought to increase the free energy of unfolding and thus the protein T_m (Niesen *et al.*, 2007). Binding of 33RE to N-domain was found to shift the T_m^0 from 69.5±0.6°C by a maximum of 11.1±0.8°C, resulting in a T_m^{max} of 80.6±0.2°C (Figure 2.26). The YR mutant T_m^0 (68.8±0.1°C) resembled that of wild-type whereas the SEDSTE and SEDSTE_YR mutants displayed slightly lower stability (T_m^0 of 66.3±1.1°C and 67.5±0.7°C, respectively). In addition, all mutants had significantly lower T_m^{max} values versus wild-type (78.7±0.2°C, 75.5±0.6°C and 73.3±0.4°C for SEDSTE, YR and SEDSTE_YR, respectively, with $P<0.01$). With the YR mutant this likely resulted from a loss in Y369 hydrogen bonding enthalpy. The decreased thermal stabilization of SEDSTE, however, suggests a decrease in the ligand binding entropy since these residues do not form any direct contacts with 33RE.

These differences in melting temperatures and thermal stabilization by 33RE highlight the importance of the S_2' subsite residues and supports the notion that the identified unique

residues indirectly and synergistically influence 33RE's behaviour by affecting the protein's solvent exposure and/or dynamics.

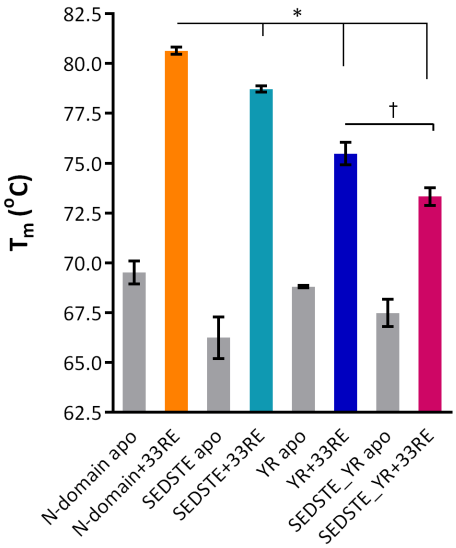


Figure 2.26: ACE thermal stability as determined by differential scanning fluorimetry.
Minimum (T_m^0) and maximum (T_m^{max}) melting temperatures are given as mean ± SD of two independent experiments in triplicate wells in the absence (grey bars) and presence (coloured bars) of 33RE, respectively. Statistical significance of mutant vs wild-type T_m^{max} was set at a 99% confidence level (*P < 0.01). † YR vs SEDSTE_YR, P = 0.052.

2.6.12 *In vitro* investigation into the structural effects of mutation

X-ray crystallography was used to evaluate the effect of active site mutation on the 33RE interactions *in vitro.* The SEDSTE_YR-33RE structure was solved by Prof K.R. Acharya and Dr G.E. Cozier (Department of Biology and Biochemistry, University of Bath). A resolution of 1.8Å was obtained for the two protein molecules in the asymmetric unit of the *P1* space group. Previously, wild-type N-domain-33RE also diffracted in the *P1* space group with two protein molecules in the asymmetric unit but the structure was solved at a slightly lower resolution of 2.2Å (Douglas *et al.*, 2014).

Crystallization trials were also done in the absence of 33RE to obtain the apo structure, potentially in an 'open' conformation. Unfortunately, none of the screened conditions allowed crystallization of the latter and the apo structure was not determined. This is potentially due to the decrease in T_m that firstly results from mutation of the N-domain active site residues (discussed in section 2.6.11) and secondly from removal of the seven PNGs (Anthony *et al.*, 2010) in creating the minimally glycosylated active site mutant. The SEDSTE_YR n389 protein is thereby probably too unstable to form stable crystal contacts. In the presence of 33RE, however, sufficient SEDSTE_YR n389 stabilization for successful crystallization is likely obtained through a ligand-induced increase in T_m (discussed in section 2.6.11), despite the effects of active site or PNGs mutation. In future, a technique such as electron microscopy might be more suitable to visualize the apo structure of a fully glycosylated protein.

Overall, the wild-type N-domain and mutant SEDSTE_YR 33RE structures were very similar and the two chains aligned with a mere 0.45Å RMSD. As with wild-type, the mutant structure showed Zn coordination via the 33RE phosphonate group, binding of the P_2' amide to Q259, K489 and Y498 (forming the oxyanion hole) and no interaction with the unique, mutated S_2' residues (Figure 2.27a). The Y369F and R381E orientations in the crystal structure were consistent with the simulated structures. In contrast to the MD results, however, mutation did not alter the orientation of 33RE and the tetrazole position was only slightly shifted (Figure 2.27b). In addition to 33RE, three ethylene glycol molecules from the crystallization solution also bound to the mutant's S_2 subsite near the 33RE tetrazole (Figure 2.27a). These ethylene glycol molecules likely masked the effects of mutation in this crystal structure by binding to the tetrazole and stabilizing its interactions to conserved residues in the absence of Y369 hydrogen bonding.

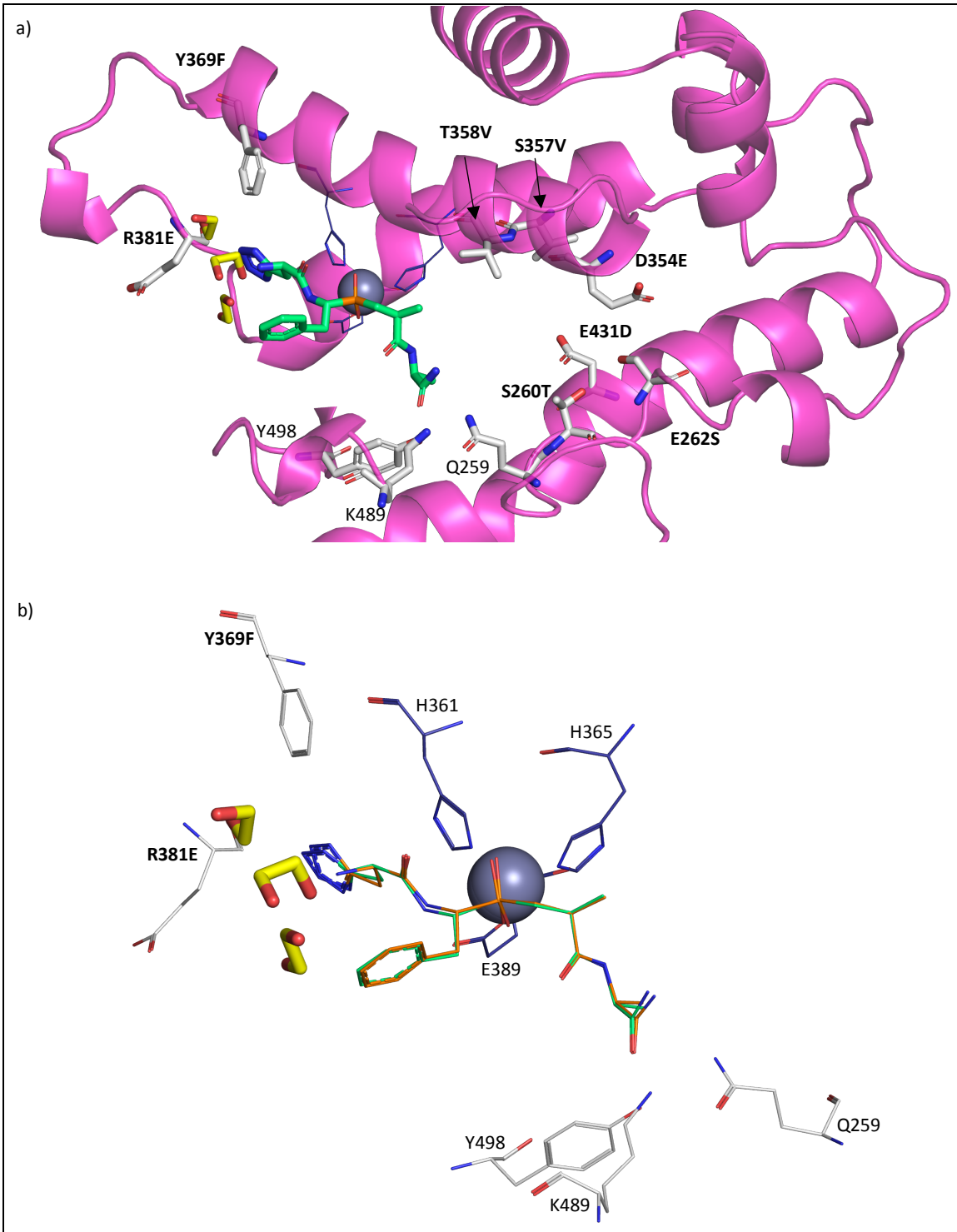


Figure 2.27: Crystal structure of the SEDSTE_YR mutant complexed to 33RE.
a) Interactions between 33RE (green sticks), ethylene glycol molecules from the crystallization solution (yellow sticks) and the mutant active site (grey sticks and pink ribbon). b) Alignment between 33re complexed to the wild-type (PDB ID: 4BXK, orange lines) and mutant (green lines) N-domain. The zinc-coordinating residues are given in blue lines and the zinc ion as a grey sphere.

The simulated separation that occurred between helices 18 and 22 upon mutation of the S_2' inter-subdomain contact point was also not observed in the crystal structure. This could be due to the nature of the contacts formed between molecules within an asymmetric unit or between neighbouring asymmetric units. Within a unit cell, numerous interactions are observed between the two molecules' subdomain 2 surface residues (Figure 2.28). Additionally, interactions are observed between unit cells via hinges 2 and 4 (molecule A) of one asymmetric unit and the lid residues (molecule A) of another (Figure 2.28). Although these contacts enable crystallization of the protein they may have masked important consequences of active site mutation by affecting the hinge flexibility and facilitating movement of the two subdomains towards each-other.

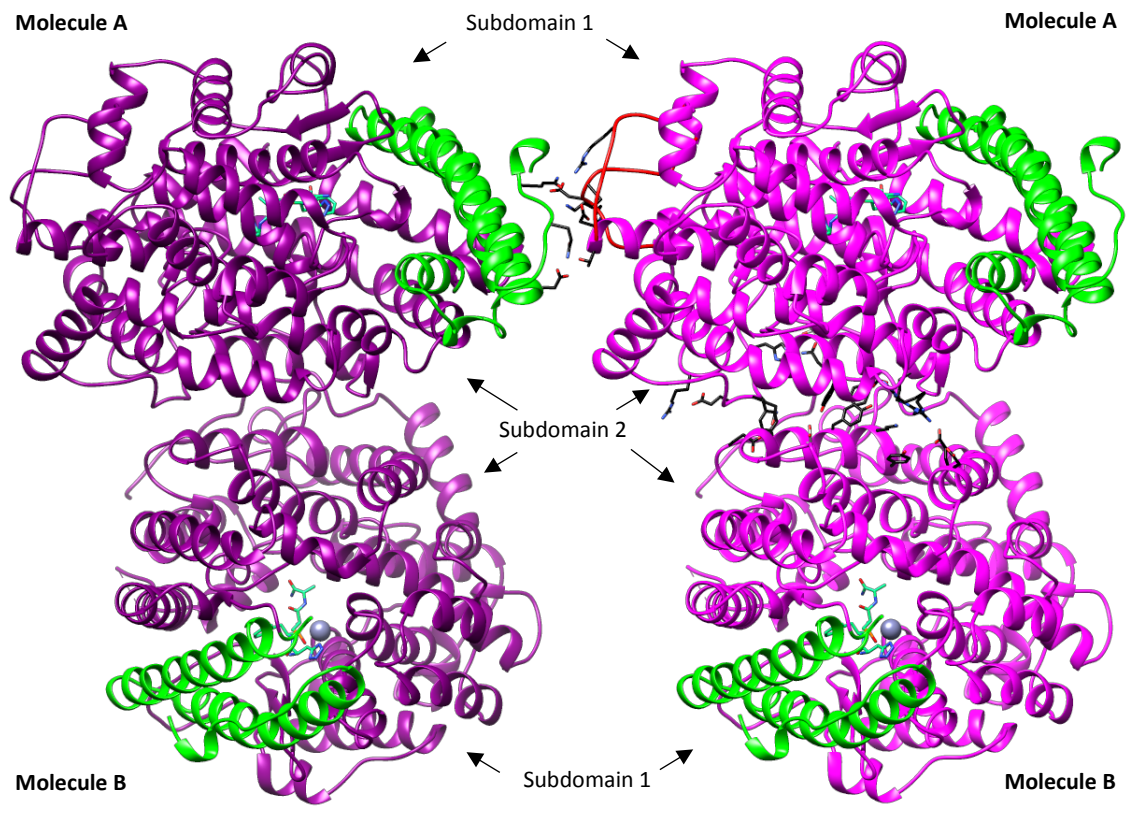


Figure 2.28: Interactions between SEDSTE_YR-33RE molecules within an asymmetric unit or between neighbouring asymmetric units.
Interactions between two molecules within the same asymmetric unit (pink) are shown on the right and interactions between molecules of neighbouring asymmetric units are shown in the upper middle (dark purple and pink Molecule A structures). Interacting residues are shown as black sticks, 33RE as green sticks, Zn as a grey sphere, the lid in green and hinges 2 and 4 in red (on Molecule A).

Mutation does, however, seem to result in protein flexibility changes *in vitro*. The alpha carbon (C_α) B-factors were compared between the two mutant or wild-type molecules present per asymmetric unit. The C_α B-factors of the two molecules in the wild-type unit cell are very similar (Figure 2.29a). In the mutant structure, however, the C_α B-factors of molecule A are increased over wild-type and that of molecule B is greatly increased over the mutant's molecule A (Figure 2.29b). This suggests that the mutant is less stable than wild-type in this 'closed' conformation. Flexibility is particularly increased at the lid and hinges 2 and 4, so much so that a part of molecule B's lid is missing (Figure 2.30). Given that the two structures were solved in the same space group and formed the same contacts between molecules, this effect on the B-factors is unlikely to be a crystal-packing artefact. The MD results showed that the two subdomains' separation is increased upon mutation of the unique S_2' residues. Since this separation is reduced to wild-type in the presence of crystal contacts, the mutated residues on helices 18 and 22 are in unwanted close contact which likely translates into increased lid and hinge flexibility.

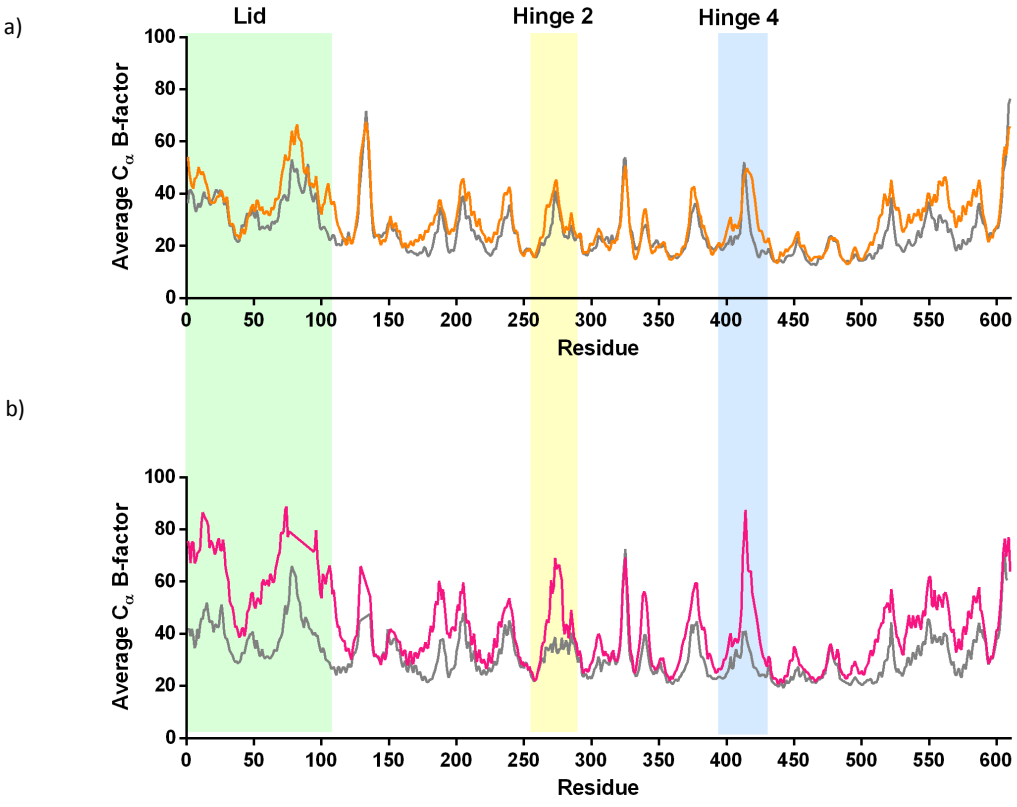


Figure 2.29: B-factor comparison of the wild-type N-domain-33RE and mutant SEDSTE_YR-33RE crystal structures.
Alpha carbon B-factors of molecule A and molecule B are respectively given in grey and in colour for a) wild-type N-domain-33re and b) mutant SEDSTE_YR-33re.

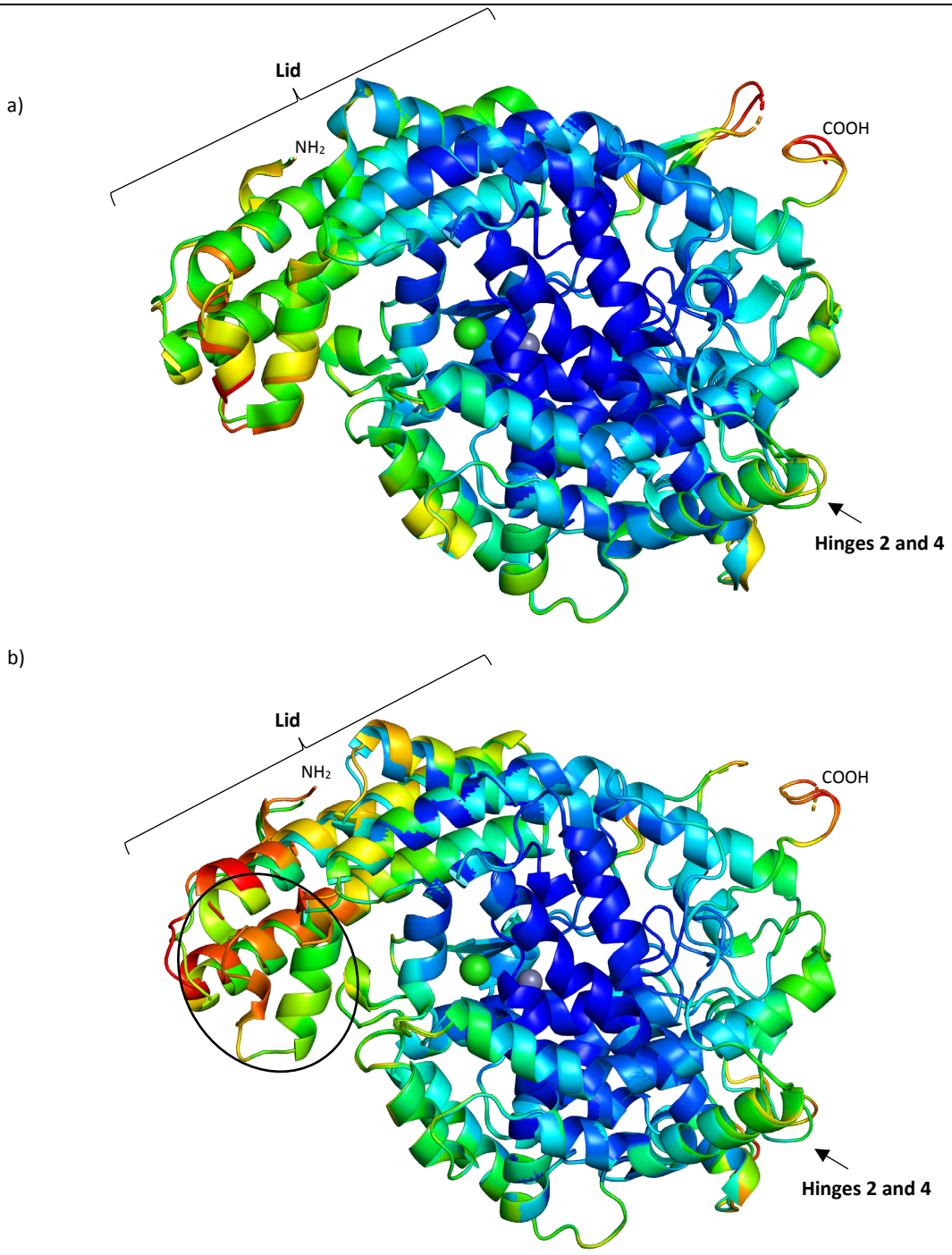


Figure 2.30: Structural view of wild-type N-domain-33RE and mutant SEDSTE_YR-33RE flexibility. The two molecules in the asymmetric unit were aligned for a) wild-type N-domain-33re (PDB ID: 4BXK) and b) mutant SEDSTE_YR-33re and coloured according to flexibility (high to low C_α B-factors from red to blue). The region of the lid that is missing from molecule B in the mutant structure is encircled. The active site Zn and chloride are given in grey and green spheres, respectively.

2.6.13 Conclusions and implications for future drug design

This study provides a clearer understanding of the mechanism whereby 33RE selectively inhibits the N-domain of ACE. Kinetic characterization of inhibitor binding highlighted the importance of the unique S_2 residues in selectivity, in line with previous studies on RXP407. Interestingly, the unique S_2' residues distal to the 33RE binding site were shown to be equally important, particularly S357 and E431. A complete loss of selectivity was seen upon combined mutation of both prime and non-prime subsites.

Although the 33RE co-crystal structure of the SEDSTE_YR mutant suggested that mutation altered protein dynamics, molecular dynamics simulations were required to elucidate the mechanism whereby these unique N-domain S_2 and S_2' residues synergistically govern 33RE's 1000-fold N-selectivity. Here we propose that the S_2 residues at hinge 3 and the hydrogen bonding network between S_2' residues facilitate active site closure upon 33RE binding through lid movement and hinging at regions 2 and 4 (Figure 2.31).

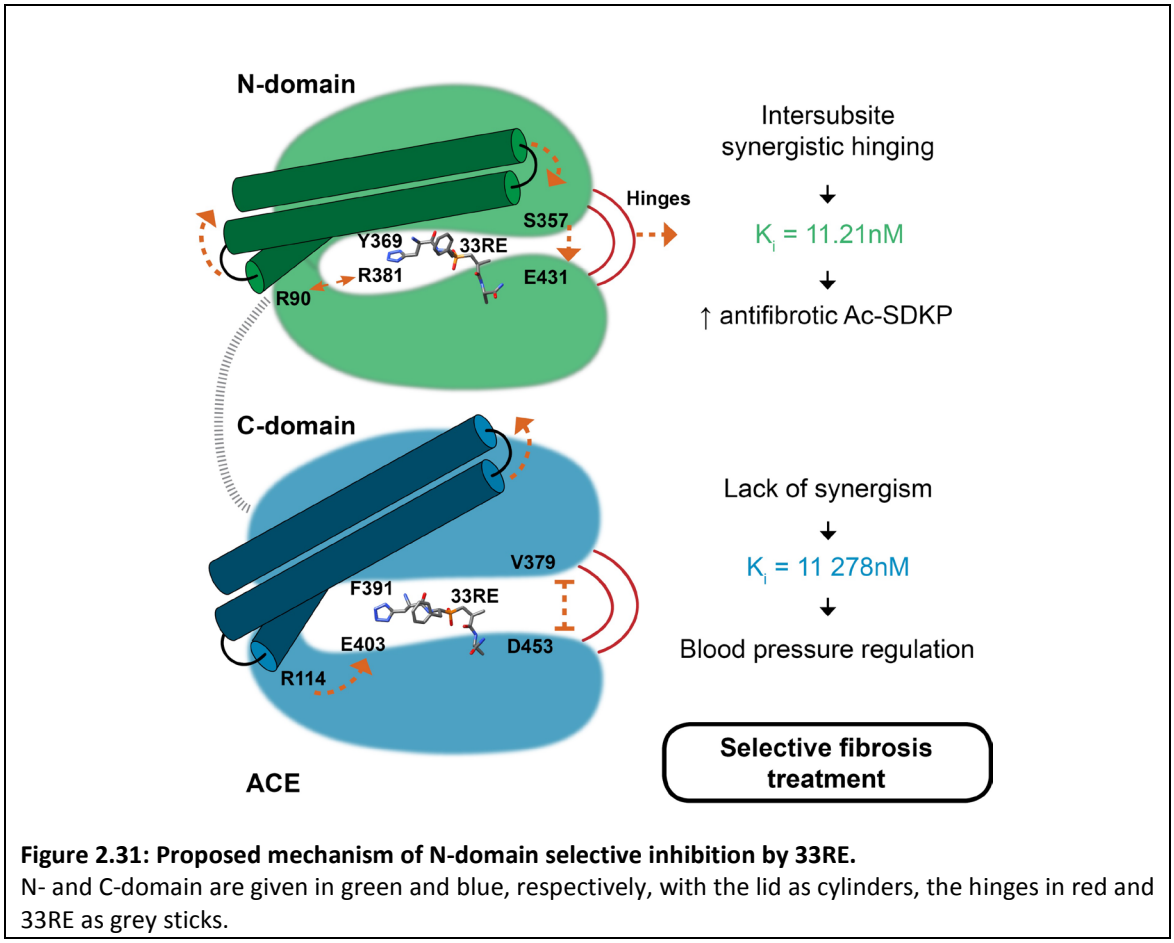


Figure 2.31: Proposed mechanism of N-domain selective inhibition by 33RE.
N- and C-domain are given in green and blue, respectively, with the lid as cylinders, the hinges in red and 33RE as grey sticks.

In accordance with the *in silico* results, differential scanning fluorimetry of fully glycosylated proteins showed a decrease in protein thermal stability and ligand thermal stabilization upon S_2' subsite mutation, suggesting altered protein dynamics and/or increased solvent exposure. The 1000-fold selectivity of 33RE is thus dependent on the P_2 tetrazole and P_2' amide interactions and is proposed to be effected through fundamental inter-domain differences in the dynamics of ACE.

Previous attempts at incorporating the RXP407 P_2 and P_2' moieties onto other ACE inhibitor backbones, however, did not yield the same order of N-selectivity. While incorporating a P_2 aspartic acid onto the slightly C-selective ketoACE inhibitor backbone increased N-domain selectivity, the P_2' amide did not and instead decreased inhibitor potency (Sharma *et al.*, 2012). This does not disprove our hypothesis, however, since the ketoACE analogues differed from RXP407 in their N-terminal protecting group, zinc binding, P_1' and/or P_2' moieties. Use of a bulky N-terminal *t*-Boc instead of an acetyl protecting group may have decreased the potency by disrupting S_2 interactions. Another factor that could have affected the potency and N-selectivity is lack of the small, hydrophobic L-pseudoalanine in the P_1', a key feature of RXP407 (Dive *et al.*, 1999).

All ketoACE analogues also possessed a carbonyl zinc binding group which would yield a strong zinc-binding geminal diol upon nucleophilic attack by a water molecule (Watermeyer *et al.*, 2008). The use of strong zinc binding groups generally tends to yield less selective metalloprotease inhibitors (Dive *et al.*, 2004). Weaker-binding phosphinic peptide inhibitors might rely more on other moieties for binding affinity so that more favourable interactions by these in the N-domain environment can bring about selective inhibition. For the highly C-domain selective RXPA380 inhibitor the phosphonate likely also allowed for selection based on the P_2 and P_2' moieties. In this case the P_2 phenylalanyl was seen to interact through Van der Waals and pi stacking with the Phe391 residue (C-domain counterpart of Y369) on subdomain 2 while the P_2' tryptophan carboxylate formed hydrogen bonds to the loop between helices 12-13 and hydrophobic interactions to the unique V379 and V380 residues (C-domain counterparts of T358 and S357 (Corradi *et al.*, 2006)). RXPA380 thus bridged the active site cleft with its binding synergistically governed by the unique C-domain residues of the S_2 and S_2' subsites (Kroger *et al.*, 2009).

Considering the RXP407, 33RE and RXPA380 results it thus appears that the right balance between prime/nonprime and subdomain 1/subdomain 2 binding is needed to attain three orders of domain-selectivity. We hypothesize that addition of an appropriate combination of subdomain 1/subdomain 2-binding moieties onto current non-selective ACE inhibitors to achieve this balance might yield N-selective compounds. Furthermore, the insight gained into the intrinsic differences in N- and C-domain dynamics could be exploited during future design of fibrosis drugs. It also offers opportunities for the design of allosteric inhibitors for targeting distal binding sites like the lid C-terminus as recently reported for the ACE homolog neurolysin (Hines *et al.*, 2014). Alternatively, modification of non-selective, bioavailable ACE inhibitors for improved interactions with the distal unique N-domain prime subsite residues through moieties unfavorable for C-domain closure might lead to the development of clinically relevant N-domain selective inhibitors.

Chapter 3

Evaluating N-selective catalytic inactivation

3.1 Introduction

3.1.1 Design of catalytic metallodrugs to target ACE

In recent years, the design of catalytic metallodrugs has received increased attention due to their potential for irreversible target enzyme inactivation at sub-saturating dosages. Design of such compounds, however, could be more challenging since the drug's target-recognition and metal-binding domains should be carefully tailored to obtain selective target inactivation through the production of metal-bound ROS with high catalytic efficiency under physiological conditions (Yu and Cowan, 2017). This can be achieved by coupling a moiety selective for the target of interest via a linker to a metal chelate, optimally oriented towards target residues with high reactivity to oxidation (Cowan, 2008). Once the metallodrug is bound to the target, the metal's effective concentration is increased and ROS production is catalytically efficient in the presence of physiological redox co-reactants such as ascorbate and H_2O_2.

Following development of the antihypertensive catalytic metallodrug $[KGHK\text{-}Cu]^+$ (Gokhale and Cowan, 2005; Gokhale and Cowan, 2006), the Cowan group designed a series of compounds specifically targeted to ACE by coupling metal chelates to the lysine sidechain of the clinically used ACE inhibitor lisinopril (Joyner *et al.*, 2012). The lysine sidechain was thus used as a linker between the target recognition (lisinopril) and metal binding domains (transition metal chelate). Chelates varying in size, charge, reduction potential and coordination unsaturation were complexed to Fe^{3+}, Co^{2+}, Cu^{2+} and Ni^{2+} transition metals (Figure 3.1)

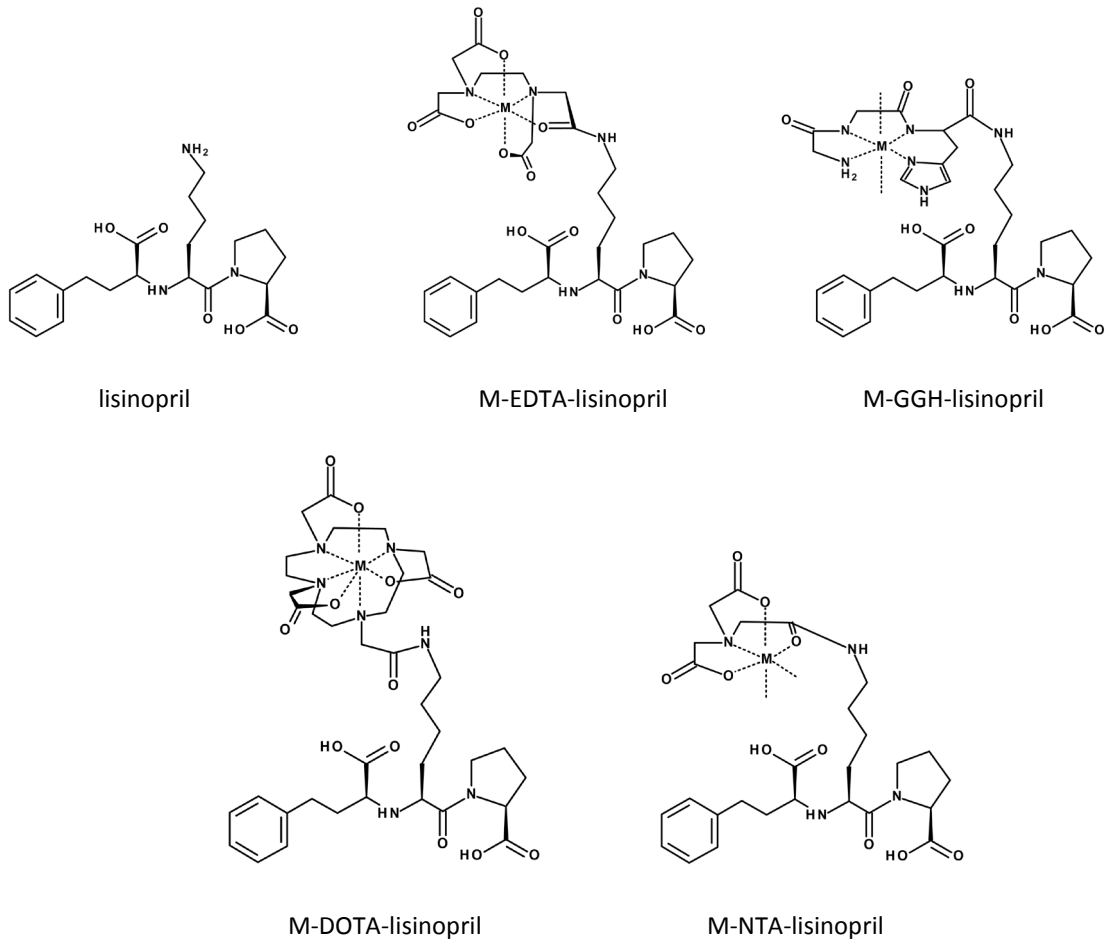


Figure 3.1: Lisinopril-derived metal chelate complexes designed by Joyner *et al* (Joyner *et al.*, 2012). M = Fe^{3+}, Co^{2+}, Ni^{2+} or Cu^{2+}.

The ability of these metallodrugs to catalyze irreversible inactivation of sACE was investigated by monitoring hydrolysis of a non-domain selective substrate. The Cu-Gly-Gly-His-lisinopril (CuGGHLis) metallodrug (Figure 3.2) was the most efficient catalyst in the presence of ascorbate and H_2O_2 and caused inactivation presumably through side-chain oxidation of amino acids involved in substrate binding and hydrolysis. A time-dependent decrease in activity was observed at a sub-stoichiometric concentration of CuGGHLis. This suggests that following inactivation, the metallodrug dissociated from the target molecule and the reduced metal state was regenerated to allow multiple turnover reactions. Importantly, this metallodrug did not produce diffuse radicals (Joyner *et al.*, 2012) (which might cause off-target inactivation) but rather metal-bound ROS, presumably an active Cu^{3+}=O^{2-} oxone intermediate (Cowan, 2008). This GGH tripeptide is part of the amino terminal copper and nickel binding (ATCUN) motifs which play important roles in the transport of Cu^{2+} and Ni^{2+} in biological systems (Harford and Sarkar, 1997; Lau and Sarkar, 1971) and are known to stabilize the $Cu^{3+/2+}$ redox couple (McDonald *et al.*, 1997).

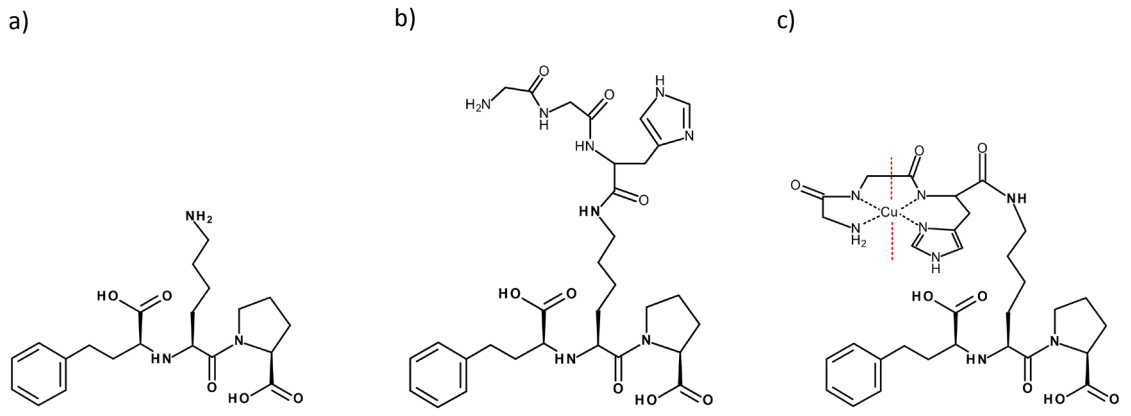


Figure 3.2: Chemical structures a) the parent compound lisinopril, b) the GGHLis metal chelate and c) the CuGGHLis metallodrug.
The two axial faces accessible for binding to target molecule are given as red lines.

3.1.2 Chemistry of the CuGGH peptide

At a pH > 6.7 where the histidyl imidazole nitrogen is deprotonated, Cu^{2+} induces deprotonation of the glycyl-glycyl peptide nitrogens (Lau *et al.*, 1974). Subsequently, quadridentate coordination of Cu^{2+} with the deprotonated peptide nitrogens, N-terminal amine and histidyl nitrogen results in a square planar complex consisting of two 5-membered and one 6-membered ring (Figure 3.2) (Camerman *et al.*, 1976). Such motifs provide the strongest peptide binding site for metal ions capable of deprotonating peptide nitrogens and the most stable complex if a histidyl residue is present in the third position (Sigel and Martin, 1982; Gonzalez *et al.*, 2018). Although Zn^{2+} might also promote deprotonation of the peptide nitrogens, it has been shown to form binuclear complexes with GGH and its lower Lewis acidity and preference for the tetrahedral geometry makes it less likely to support the square planar quadridentate form at a pH < 9 (Agarwal and Perrin, 1977; Gonzalez *et al.*, 2018). Ligand exchange of CuGGHLis with Zn^{2+} present in the active site of ACE is thus not expected to occur. Indeed, the absence of ligand exchange has been demonstrated experimentally for CuGGHLis in the presence of a molar equivalent of Zn^{2+} (Hocharoen *et al.*, 2013).

To be suitable for physiological use, the metallodrug needs to bind very tightly to the metal and the resulting complex must be stable to avoid loss of the active metallodrug and generation of off-target effects through depletion of essential metal ions by the free chelate. Metallodrugs incorporating ATCUN peptides are unlikely to cause toxicity or side-effects since ATCUN motifs

are naturally present in biological systems at the N-terminus of serum albumins (Harford and Sarkar, 1997). Furthermore, GGH binds Cu^{2+} with a very high affinity (K_D = 1.18 x 10^{-16} M for Cu^{2+}) (Lau *et al.*, 1974). Upon administration, CuGGHLis would thus be resistant to Cu^{2+} extraction by other metal-sequestering agents and catalytic efficiency would be maintained.

3.1.3 Domain-selectivity of lisinopril metal chelates

Interestingly, by monitoring sACE activity in the presence of metallodrug and redox co-reactants with domain-selective substrates, Hocharoen *et al* found that CuGGHLis selectively catalysed irreversible inactivation of the N-domain although it reversibly inhibited both domains (Hocharoen *et al.*, 2013). It was proposed that the CuGGH moiety might be more optimally orientated in the N-domain than C-domain towards sidechain oxidation of critical amino acid residues but the exact mechanism of inactivation was not elucidated since the commercial sACE enzyme's concentration was too low for mass spectrometry experiments.

The recombinant human sACE used in their study was originally obtained from an NS0-derived murine myeloma cell line (R&D systems) (Hocharoen *et al.*, 2013). No decrease in sACE activity was reported in the presence of redox co-reactants but absence of CuGGHLis. Previously, however, Chen *et al* observed that sACE expressed in cultured endothelial cells was completely inactivated within 4 hours by diffuse radicals generated by phorbol 12-myristate 13-acetate (PMA)-activated neutrophils (Chen *et al.*, 1992; Chen and Catravas, 1993). Oxidative inactivation of purified sACE was also described by Voronov *et al* and Michel *et al* and the C-domain was shown to be more susceptible than the N-domain to diffuse radicals (Voronov *et al.*, 2002b; Voronov *et al.*, 2003; Michel *et al.*, 2001). Since protein glycosylation varies greatly between host cell systems and susceptibility to oxidation by diffuse radicals is dependent on protein solvent accessibility, the disparities might have resulted from use of NS0-cell derived sACE (Li *et al.*, 2017; Goh and Ng, 2017). NS0 cell-expressed immunoglobulin G (IgG) has been shown to have a higher abundance of sialylated and high-mannose glycans than IgG found in human plasma while that expressed in CHO-K1 cells was more comparable to human (Goh and Ng, 2017).

Although the use of domain-selective ACE substrates allowed screening of the metallodrug series and identification of the most promising domain-selective inactivator, the absolute binding affinity and inactivation rate may have been influenced by two factors. Firstly, the

domain-selective ACE substrates Abz-LFK(Dnp)-OH and Abz-SDK(Dnp)P-OH used only display moderate domain-selectivities of approximately 75-fold C- and 50-fold N-selectivity, respectively (Rousseau *et al.*, 1995; Bersanetti *et al.*, 2004). Furthermore, this results from differences in the rate of catalysis of each substrate by the two domains and not a difference in their binding affinity. Since Abz-SDK(Dnp)P-OH binds equally to both domains but its catalysis is only 50-fold slower by the C-domain, complete inhibition would only be observed once CuGGHLis (in the absence of redox co-reactants), for example, has out-competed the substrate from both the N- and C-domain. The individual domains' inhibitor binding affinities might thus be underestimated when evaluating full-length sACE activity with these substrates. Secondly, these underestimated inhibitor binding affinities might have caused an overestimation of the initial rate of time-dependent oxidative inactivation. The inactivation experiments were performed in the presence of redox co-reactants and metallodrug at its IC_{20} (thus IC_{50} x 2/5) (Hocharoen *et al.*, 2013). CuGGHLis, for example, was present at 600nM or 60nM when monitoring N- or C-domain activity, respectively. The observed initial rate of inactivation depends on the inhibitor and enzyme concentrations and thus cannot be directly compared between the two domains. Since the second order rate constants which would account for these differences were not reported, the selectivity of CuGGHLis-mediated catalytic inactivation is uncertain at present. Moreover, it is not clear whether the observed rapid decrease in Abz-SDK(Dnp)P-OH activity is due to a combination of discreet N-domain oxidation and extensive C-domain oxidation since the latter is completely saturated (based on K_i determined using Abz-LFK(Dnp)-OH) by the 600nM CuGGHLis used in this reaction.

3.2 Aims and objectives

The research described in this chapter aimed to critically evaluate the binding affinity and catalytic efficiency of CuGGHLis towards reversible inhibition and irreversible catalytic inactivation of the two ACE domains and to gain insight into the mechanism whereby this occurs.

The objectives to achieve this aim were as follows:

1. Characterize the affinity of lisinopril, GGHLis and CuGGHLis binding to N- and C-domain by:
    a. Obtaining truncated forms of the N- and C-domain proteins by affinity purification following transfection and heterologous protein expression in a CHO-K1 mammalian cell system

    b. Evaluating inhibitor binding affinity to each domain in isolation using a non-domain selective fluorogenic substrate

2. Evaluate the domain-selectivity of CuGGHLis-mediated oxidative inactivation by:
    a. Monitoring the decrease in N- or C-domain activity over time in the presence of metallodrug and redox co-reactants using a non-domain selective fluorogenic substrate and determining the initial rates and second order rate constants for inactivation

3. Gain insight into the mechanism whereby oxidative inactivation occurs by:
    a. Characterizing the residual activity after inactivation by determining Michaelis-Menten kinetic parameters for hydrolysis of a non-domain selective fluorogenic substrate

    b. Using SDS-PAGE to determine whether inactivation is as a result of amino acid sidechain oxidation or protein backbone cleavage

3.3 Methodology

Deionized MilliQ® distilled water with a resistivity of 18.2 MΩ.cm (at 25°C) was used in the preparation of all solutions throughout this chapter (except tissue culture media and buffers). Buffers (containing no detergent) were filter-sterilized, stored at 4°C and warmed to ambient temperature prior to use.

3.3.1 Preparation of metallodrug

3.3.1.1 Preparation of a $CuCl_2$ solution

To obtain a copper solution of known concentration for Gly-Gly-His-Lisinopril (GGHLis) titration, volumetric standardization was employed. Copper chloride dihydrate (Fluka™) was dissolved in deionized water before dropwise dilution with a titration buffer (20mM HEPES pH 7.5 and 100mM NaCl) to a final concentration of 3mM $CuCl_2$ while mixing. The Cu^{2+} concentration of this solution was verified by titration into a solution of 4mM EDTA (Merck) prepared in the aforementioned titration buffer. This was performed in the absence of zinc to prevent formation of a ZnEDTA complex.

3.3.1.2 Titration of EDTA with Cu

Titration buffer to a volume of 450µl was added to a quartz cuvette and the absorbance zeroed over a 190 - 800nm wavelength range using a Cary Varian 50 Conc UV-Vis spectrophotometer. Fifty microliters of the prepared 4mM EDTA was added to the cuvette while stirring and the absorbance read to establish a baseline. While constantly stirring, copper was titrated into the EDTA solution in 10µl volume intervals and the absorbance read after a 15-minute room temperature incubation at each interval. Formation of the CuEDTA complex was monitored as an increase in absorbance at a wavelength of 270nm with extinction coefficient of $3300 M^{-1} cm^{-1}$ as described by Dr J.C. Joyner (Joyner, 2012). Copper addition proceeded until no further increase was observed at this wavelength. The data was fitted to a segmental linear regression model using GraphPad Prism® v6.01. The prepared stock solution's Cu^{2+} concentration was subsequently determined from the molar equivalent of copper to EDTA at the inflection point on the curve.

3.3.1.3 Titration of GGHLis with Cu

Lyophilized GGHLis was a kind gift from Prof J.A. Cowan at Ohio State University. It was dissolved in deionized distilled water to a stock concentration of 7.4mM and stored at -20°C.

The CuGGHLis metallodrug was prepared by titration with the EDTA-standardized $CuCl_2$ solution. This was performed in the absence of zinc to prevent formation of a ZnGGHLis complex. Titration buffer (20mM HEPES pH 7.5 and 100mM NaCl) to a volume of 450µl was added to a quartz cuvette and the absorbance zeroed over a 190 - 800nm wavelength range using a Cary Varian 50 Conc UV-Vis spectrophotometer. Following equivolume dilution of the GGHLis chelator into the titration buffer, 50µl GGHLis was added to the cuvette while stirring. The GGHLis concentration was confirmed by an absorbance peak at 257nm with extinction coefficient of $197 M^{-1} cm^{-1}$ due to presence of the lisinopril phenylalanine moiety.

While constantly stirring, copper was titrated into the GGHLis solution in 10µl volume intervals and the absorbance read after a 15-minute room temperature incubation at each interval. Formation of the CuGGHLis complex was monitored as a shift in the 257nm peak and an increase in absorbance at a wavelength of 250nm with extinction coefficient of $4180\ M^{-1} cm^{-1}$ as described by Dr J.C. Joyner (Joyner, 2012). Detection of the d-d transition at 525nm (Lau *et al.*, 1974) was less pronounced and thus not used for complex quantification. Copper addition proceeded until no further increase was observed at 250nm. Absorbance data was fitted to a segmental linear regression model using GraphPad Prism® v6.01 and the titration end-point determined from the inflection point.

Two additional copper titrations were performed for GGHLis. The second titration was performed up until the established endpoint to obtain a 1:1 molar ratio of metal to chelator. The metallodrug was aliquoted in 20µl volumes, stored at -20°C and used to characterize inhibitor binding. The third titration was halted prior to the endpoint at a 1:1.1 metal to chelator molar ratio. The metallodrug was aliquoted in 20µl volumes, stored at -20°C and used to characterize catalytic inactivation. The slight excess of chelate ensured that no free copper ions were present and that only metallodrug-bound reactive oxygen species would be produced. The lack of free copper ions with these two titrations was further confirmed by the absence of absorbance at 750 – 800nm.

3.3.2 Protein expression and purification

Chinese Hamster Ovary cells (CHO-K1) previously stably transfected with fully glycosylated N- or C-domain (Kroger *et al.*, 2009) were reconstituted into growth medium. They were grown to confluency in T-175cm^2 flasks (Nunc Ltd.) before replacing the growth medium with harvesting medium (constituents in Appendix A6). Culture medium was harvested every 48 hours and the pooled harvests stored at -20°C until protein purification. Protein purification was performed as described in Chapter 2 with the exception that the dialysis solutions were prepared using deionized distilled water. The dialyzed and concentrated proteins were stored at -20°C in 20µl aliquots to minimize freeze-thawing and possible protein oxidation or degradation during storage. ACE activity was assessed during protein expression and purification using the non-domain selective fluorogenic substrate Z-phenylalanylhistidylleucine (Z-FHL) (Sigma-Aldrich® Co.) as described in Chapter 2.

3.3.3 Optimization of fluorogenic ACE assay

Coordination of Cu^{2+} to the GGHLis chelator and formation of the square planar complex involves deprotonation of the backbone amides and histidine ionization (Camerman *et al.*, 1976) (Figure 3.2). A pH greater than 6.5 therefore should be maintained to prevent dissociation of the copper ion. A time-dependent decrease in sACE activity upon incubation at 37°C was described by Hocharoen *et al* (Hocharoen *et al.*, 2013). This was circumvented by addition of the Brij-35 detergent at a concentration of 0.05%.

3.3.3.1 Optimization of detergent concentration

The optimal Brij-35 concentration required for stability of the truncated N- and C-domain proteins was determined by incubation of the proteins at 4°C or 37°C across a range of detergent concentrations. A 1% (w/v) Brij-35 (Merck) solution was prepared in deionized distilled water and stored at room temperature. Dilutions of Brij-35 into an assay buffer consisting of 50mM HEPES pH 8.2, 300mM NaCl and 10µM $ZnCl_2$ were freshly prepared from this stock solution each day.

The non-selective ACE substrate Hippuryl-His-Leu (HHL) was prepared by dissolving 24.5mg HHL (Sigma-Aldrich® Co.) in 2.083ml NaOH (0.025M) with heating to 40°C and vortexing according to the protocol described by Schwager *et al* (Schwager *et al.*, 2006). Assay buffer consisting of

50mM HEPES pH 8.2, 300mM NaCl and 10µM $ZnCl_2$ was added dropwise to yield a 5.7mM stock solution. Solutions of 2mM HHL, 7nM N-domain and 2nM C-domain were subsequently prepared in assay buffer containing Brij-35 ranging from 0.0004% to 0.5%. The enzymes were incubated at 4°C or 37°C for four hours after which 20µl was aliquoted in triplicate wells of a 96-well plate. Twenty microliters of 2mM HHL was added to each well and the plate incubated at 37°C for 15 minutes for substrate hydrolysis to occur. This was followed by termination of the reaction with 177µl 0.28M NaOH and addition of 16µl O-phthaldialdehyde (20mg/ml (Sigma-Aldrich® Co.)) in methanol. Incubation for 10 minutes at room temperature with shaking allowed derivatization of the HL sequestered by Z-FHL hydrolysis. The reaction was stopped by addition of 30µl HCl (3M) and the fluorescence intensity detected at excitation and emission wavelengths of 360nm and 485nm, respectively, using a fluorescence spectrophotometer (Cary Eclipse Varian Inc.). The assay protocol followed was analogous to that described by Schwager *et al* (Schwager *et al.*, 2006). Fluorescence intensity was plotted against detergent concentration and the Brij-35 concentration yielding optimal stabilization determined. Stability at this concentration was confirmed by incubating the enzymes (in assay buffer containing 0.0015% Brij-35) at 22°C or 37°C for 100 minutes and assaying their activity at 20-minute intervals.

3.3.4 Characterization of substrate hydrolysis:

3.3.4.1 Construction of a standard curve

A standard curve was constructed using His-Leu (HL) standards (Sigma-Aldrich® Co.) prepared in assay buffer (50mM HEPES pH 8.2, 300mM NaCl, 10µM $ZnCl_2$ and 0.0015% Brij-35) to relate fluorescence intensity to moles of HL, and thus enzyme activity (Appendix A14).

3.3.4.2 Kinetic characterization of HHL hydrolysis

The amount of active protein was determined by calculating the specific activity from Z-FHL (Schwager *et al.*, 2006) and Bradford (Bradford, 1976) assays as described in Chapter 2.

Binding and hydrolysis of the non-domain selective HHL substrate was determined in an assay buffer consisting of 50mM HEPES pH 8.2, 300mM NaCl, 10µM $ZnCl_2$ and 0.0015% Brij-35. Purified N- or C-domain was diluted in this assay buffer to 15nM or 4nM, respectively. An HHL dilution series ranging from 0.04mM to 2.85mM was also prepared in assay buffer. Enzyme and substrate were pre-warmed to 37°C for 5 minutes before 5µl enzyme and 28µl substrate was

aliquoted to triplicate wells of a 96-well plate. The plate was incubated for 15 minutes at 37°C on an orbital shaker to allow substrate hydrolysis and the rest of the procedure described in section 3.3.3.1 followed.

The resulting velocity data was fitted to a non-linear Michaelis-Menten model using GraphPad Prism® v6.01. The parameters describing HHL binding affinity (K_m), turnover (k_{cat}) and catalytic efficiency (k_{cat}/K_m) were subsequently determined for each domain from duplicate independent assays.

3.3.5 Kinetic characterization of inhibitor binding affinity

A 10mM stock solution of lisinopril (Sigma-Aldrich® Co.) was prepared in deionized distilled water, aliquoted and stored in the dark at room temperature. Preparation of GGHLis and CuGGHLis was described in section 3.3.1.3. A dilution series of each inhibitor was prepared in assay buffer consisting of 50mM HEPES pH 8.2, 300mM NaCl, 10µM $ZnCl_2$ and 0.0015% Brij-35. The CuGGHLis used in this section was prepared at a 1:1 molar ratio of metal to chelator. Solutions of 15nM N-domain, 4nM C-domain and 2mM HHL were also prepared in assay buffer. Forty microliters of enzyme was incubated with 40µl of inhibitor at each concentration for 20 minutes at 22°C and the substrate pre-warmed to 37°C. Twenty microliters of the enzyme-inhibitor mixture was aliquoted in triplicate wells of a 96-well plate before addition of 20µl HHL (2mM) to each well. The plate was incubated for 15 minutes at 37°C on an orbital shaker to allow substrate hydrolysis and the rest of the procedure described in section 3.3.3.1 followed.

Activity at each inhibitor concentration was calculated as a percentage relative to the uninhibited control and fitted to a sigmoidal dose-response curve using GraphPad Prism® v6.01. The inhibitor concentration at which 50% enzyme inhibition occurred (IC_{50}) was determined and the inhibitor binding affinity (K_i) calculated using the Cheng-Prusoff equation (Cheng and Prusoff, 1973) (equation 2) where [S] = [HHL] = 1mM (final concentration) and K_m determined for each domain taken from the results of section 3.3.4.2.

$$K_i = \frac{IC_{50}}{\left(1 + \frac{[S]}{K_m}\right)} \tag{2}$$

An alternative method of determining K_i values is via Dixon plots (as described in Chapter 2). For CuGGHLis, however, not only the K_i but also the IC_{50} values were required for downstream experiments. The Cheng-Prusoff equation allowed determination of both parameters from a single data set.

3.3.6 Determination of the mode of inhibition

To investigate whether CuGGHLis competed with substrate for binding to the active sites of the truncated single-domain proteins, inhibition was assessed using the Dixon method (Dixon, 1953). This involved measurement of enzyme activity at constant enzyme but two different substrate concentrations and a range of inhibitor concentrations. The procedure described in section 3.3.5 was followed. Duplicate samples were incubated with CuGGHLis (at a 1:1 molar ratio of metal to chelator) and each set assayed with either 20µl of 2mM HHL or 20µl of 1mM HHL per well. The inverse fluorescence was plotted as a function of inhibitor concentration at two different substrate concentrations to yield a Dixon plot of which the shape revealed the mode of inhibition.

3.3.7 Time-dependent oxidative inactivation

3.3.7.1 Preparation of redox co-reactants

The redox co-reactants ascorbate and H_2O_2 were made fresh each day immediately prior to use and kept on ice and in the dark. The stock solution of stabilized H_2O_2 (30%) (Sigma-Aldrich® Co.) (stored at 4 degrees) was diluted to 40mM in an assay buffer consisting of 50mM HEPES pH 8.2, 300mM NaCl, 10µM $ZnCl_2$ and 0.0015% Brij-35. One gram of L-ascorbic acid (Sigma-Aldrich® Co.) was dissolved in assay buffer to a concentration of 100mM and the pH adjusted to 8.2 using NaOH pellets. The pH meter was calibrated each day using commercial standards (standard solutions of pH 4.00 and pH 10.00 Sigma-Aldrich® Co.). Further dilution of ascorbate into assay buffer yielded a 40mM solution. A solution containing both co-reactants at 40mM was also prepared.

3.3.7.2 Time-course inactivation procedure

Solutions of 15nM N-domain, 4nM C-domain and 2mM HHL were freshly prepared in assay buffer and the enzyme kept on ice. The CuGGHLis metallodrug prepared at a 1:1.1 molar ratio

of metal to chelator was used in this section. It was diluted in assay buffer to a concentration of $2xIC_{20}$ for each domain. The IC_{20} values were calculated from the IC_{50} values obtained for N- and C-domain in section 3.3.5 as $IC_{20} = IC_{50}*2/5$. One thousand six hundred microliters enzyme and 1600µl metallodrug or buffer were incubated for 20 minutes at 22°C in the absence of redox co-reactants. Prior to oxidative inactivation, each sample containing CuGGHLis retained 80% activity. Two microliter aliquots of buffer, ascorbate, H_2O_2 or both co-reactants (each at 0.04M) were aliquoted to separate PCR tubes for a range of time-points for each sample and kept on ice. Eighty microliters of the enzyme-inhibitor or enzyme-buffer reactions were aliquoted to each tube and incubated for 8 hours at 22°C to allow oxidative inactivation. The final co-reactant concentrations were 1mM. The use of PCR tubes ensured that no sample loss occurred due to evaporation over this prolonged time-period. Timing of the inactivation reaction commenced upon first contact of the enzyme with co-reactant. Samples were immediately assayed for activity at each timepoint using 2mM HHL as described in section 3.3.5. Activity at each time-point was expressed as a percentage relative to the control sample (enzyme containing no co-reactant or metallodrug) as a function of time. The data was fitted to a non-linear exponential decay model using GraphPad Prism® v6.01.

3.3.7.3 Kinetic characterization of residual activity

To gain insight into the mechanism of catalytic inactivation, the activity remaining was kinetically characterized. The 100-minute inactivation timepoint was chosen to allow a degree of inactivation but still retain enough residual activity for detection with an HHL assay. An HHL dilution series ranging from 0.04mM to 2.85mM was prepared in assay buffer. Solutions of 44nM N-domain or 19nM C-domain (prepared in assay buffer) were incubated for 20 minutes at 22°C with equivolume amounts of buffer or CuGGHLis at IC_{20}. These reactions were aliquoted to tubes containing buffer or co-reactants (at 40mM) and incubated for 100 minutes at 22°C to allow protein oxidation. The final co-reactant concentrations were 1mM. The samples were put on ice and 5µl aliquots of each sample added to triplicate wells of a 96-well plate. Twenty-eight microliters of HHL was added to each sample and the activity assayed across the range of substrate concentrations. The data was fitted to the non-linear Michaelis-Menten model using GraphPad Prism® v6.01 for determination of the K_m and k_{cat} parameters.

3.3.7.4 Assessment of artificial protease activity of CuGGHLis

To determine whether any observed time-dependent decrease in activity was due to CuGGHLis-mediated oxidative protein cleavage, the samples were analysed by SDS-PAGE. A master dilution of 1.28µM N- or C-domain and CuGGHLis solutions of $2xIC_{20}$ were prepared in assay buffer and enzyme incubated with buffer/metallodrug for 20 minutes at 22°C. The enzyme-metallodrug or enzyme-buffer reactions were aliquoted to tubes containing buffer or co-reactants (to a final concentration of 1mM) and incubated for 7 hours at 22°C to allow protein oxidation. A 15mM stock solution of EDTA was prepared in 50mM HEPES pH 8.0. The inactivation reactions were quenched at the 7-hour timepoint by the addition of EDTA to a final concentration of 1.5mM. The samples were reduced, denatured and resolved using a 10% polyacrylamide gel with Colloidal Blue-Silver Coomassie staining (Appendix A15) to detect oxidative cleavage fragments. A control sample which was not incubated for 7 hours and contained no co-reactants or metallodrug was included as a control. Each sample contained 500ng of protein.

3.4 Results and discussion

3.4.1 Standardization of a $CuCl_2$ solution

To obtain a copper solution of known concentration for titration of Gly-Gly-His-Lisinopril (GGHLis), volumetric standardization was employed. An EDTA solution of known concentration was titrated with $CuCl_2$ and formation of the CuEDTA complex spectrophotometrically monitored (Appendix A13 Figures A6 and A7).

3.4.2 Preparation of the CuGGHLis metallodrug

The CuGGHLis metallodrug was prepared by titration with the EDTA-standardized $CuCl_2$ solution. The CuGGHLis titration end-point was determined (Appendix A13 Figures A8 and A9) before preparing the metallodrug to a 1:1 molar ratio of metal to chelate for binding affinity characterization (Appendix A13 Figure A10). The third titration of GGHLis was halted prior to the endpoint at a 1:1.1 molar ratio of metal to chelate. This ensured that no free copper ions would be present during the oxidative inactivation experiments.

3.4.3 Protein purification

The truncated N- and C-domain constructs were stably transfected into CHO-K1 cells and the proteins purified using affinity chromatography. Two rounds of dialysis were performed in buffer prepared using deionized distilled water to ensure that no metal ions were present. SDS-PAGE and Coomassie staining revealed bands of ~90kDa for N-domain and ~70kDa for C-domain (Figure 3.7). The difference in molecular weight between the domains is indicative of the greater extent of N-domain glycosylation (nine N-domain compared to six C-domain PNGs as described in Chapter 1 section 1.3).

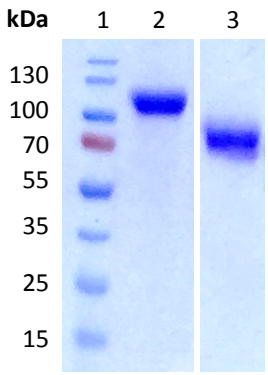


Figure 3.7: Coomassie-stained SDS-PAGE gel (10%) indicating protein size and purity.
Lane 1: protein ladder (kDa), lane 2: N-domain and lane 3: C-domain.

3.4.4 Optimization of a fluorogenic ACE assay

The assay buffer used for the Z-FHL assay described in Chapter 2 consisted of 100mM K_2HPO_4 /KH_2PO_4 pH 8.3, 300mM NaCl and 10µM $ZnCl_2$. Copper chloride used in preparation of the CuGGHLis metallodrug, however, precipitated in this buffer. The titrations were therefore performed in a HEPES buffer. To maintain consistency in the buffers used in this Chapter, the ACE activity assay was optimized for a HEPES buffer system. The HEPES assay buffer used by Hocharoen *et al* consisted of 50mM HEPES pH 7.4, 300mM $ZnCl_2$ and 10µM $ZnCl_2$ (Hocharoen *et al.*, 2013). These conditions, however, resulted in instantaneous precipitation of the Z-FHL as well as HHL substrate. Precipitation of the HHL substrate was prevented by increasing the pH to 8.2. Since the Z-FHL substrate precipitated out at this pH after approximately one hour at room temperature, all enzyme assays from this point onwards were performed using the HHL substrate.

It was previously described that sACE undergoes a time-dependent decrease in activity upon incubation at 37°C (Hocharoen *et al.*, 2013). This would pose a problem for the assessment of CuGGHLis-mediated catalytic inactivation over extended time-periods. For sACE, this was circumvented by the addition of 0.05% Brij-35 detergent. To confirm the truncated domains' stability at 37°C, the proteins were incubated in an assay buffer containing 0.05% Brij-35 and their HHL activity assessed at 15- to 25-minute intervals. Although the N-domain activity was largely unaffected even after two hours, the C-domain was not sufficiently stabilized by 0.05% Brij-35 and its activity decreased linearly with time (Figure 3.8).

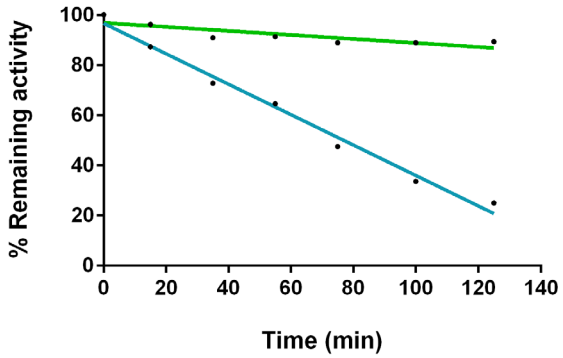


Figure 3.8: Enzyme stability at 37°C in the absence of inhibitor and co-reactants.
N-domain and C-domain are indicated in green and blue, respectively, as % remaining activity relative to time zero in a buffer containing 0.05% Brij-35 using the HHL substrate.

An in-depth investigation into the mechanism whereby Brij-35 stabilizes sACE has not been reported. Each molecule of this non-ionic surfactant consists of a hydrophobic lauryl alcohol moiety and a hydrophilic moiety consisting of 23 units of ethylene oxide. It is generally thought that protein aggregation is prevented and a protein's hydrophobic sites protected from unfavorable interactions with water molecules by the lauryl alcohol moiety's interaction with these sites, while the polyethylene oxide can increase its solubility (Otzen, 2011).

A concentration of 0.05% Brij-35 is also generally recommended to maintain matrix metalloproteinase (MMP) activity. Park *et al*, however, demonstrated that MMP-26 activity was dependent on the concentration of Brij-35 used (Park *et al.*, 2010). MMP-26 activation was observed at Brij-35 concentrations below the critical micelle concentration (CMC) whereas higher concentrations, such as 0.05%, counterintuitively resulted in non-competitive inhibition. The CMC is a property of each surfactant describing the point below which no micelles exist and above which all additional detergent molecules will form a micelle (Helenius *et al.*, 1979).

To evaluate whether the truncated domains of ACE display a similar concentration-dependence for Brij-35 stabilization, substrate and enzyme dilutions were prepared in assay buffer of varying Brij-35 detergent concentrations. Duplicate enzyme samples across the dilution series were incubated for 4 hours at either 37°C or 4°C and assayed for HHL hydrolysis. Activity of both domains was barely detectable after 4 hours at either temperature in the absence of detergent (Figure 3.9). The N-domain was stable at both temperatures and showed maximum stabilization at 0.0015% Brij-35 which was equal to that observed at 0.05% Brij-35. The maximum C-domain

stabilization was also achieved at 0.0015% Brij-35 but the activity at 37^{o}C, however, was approximately halved (vs. 4^{o}C). This emphasizes the inferior thermal stability of the C-domain which was described in Chapter 1 section 1.3.3. Contrary to the N-domain, Brij-35 concentrations greater than 0.0015% led to a decrease in C-domain activity. The observed time-dependent decrease in C-domain activity in the presence of 0.05% Brij-35 (Figure 3.8) likely occurred since it only provides a fraction of the maximum stabilization.

The mechanisms whereby Brij-35 stabilizes the N-domain, C-domain and sACE thus appear to be distinct. The optimal concentration of 0.0015% Brij-35 is below its reported CMC of 0.01% (Linke, 2009; Hait and Moulik, 2001). It is therefore proposed that the truncated domains of ACE are protected against thermal denaturation and aggregation by binding of the monomer surfactant molecules' lauryl alcohol moieties to their hydrophobic sites. Based on the biophysical results obtained in Chapter 2, it was proposed that the C-domain active site cleft is more exposed to solvent. It also has a lower *N*-glycan content than the N-domain (Chapter 1 section 1.3.1). Upon Brij-35 addition past the CMC, the large micelle structures likely interact with the more accessible C-domain but not the N-domain and decrease the stabilizing effect provided by the monomer surfactant molecules.

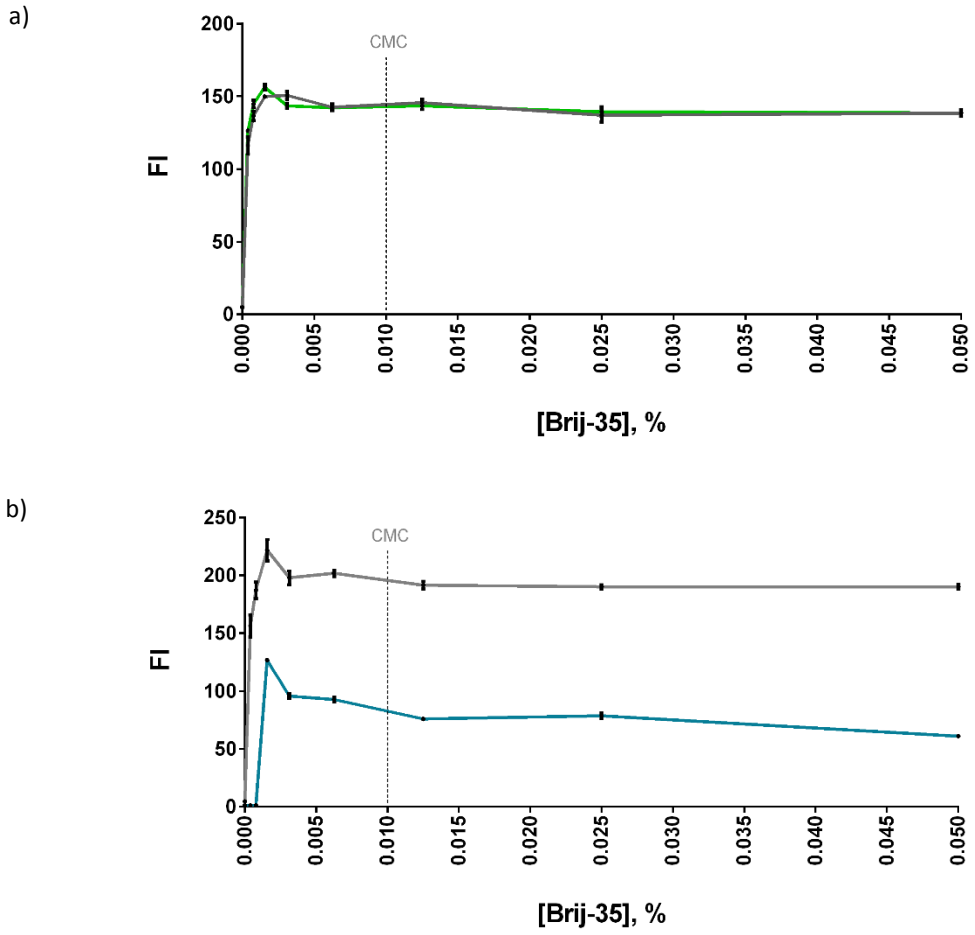


Figure 3.9: Enzyme stability after 4h incubation across a range of Brij-35 detergent concentrations. a) N-domain and b) C-domain were incubated at 4°C (grey) or 37°C (green and blue, respectively). FI: fluorescence intensity and CMC: critical micelle concentration.

The N- and C-domain stability over time was confirmed by incubation in the presence of 0.0015% Brij-35 and assessment of all time-points' HHL activity simultaneously using an inverse time-course. The experiment was performed in parallel at either 22°C or 37°C to determine whether a lower temperature would be more favourable for the less thermostable C-domain. N-domain activity was sustained for 100 minutes in the presence of Brij-35 at either temperature (Figure 3.10 a-b). The C-domain displayed a slight time-dependent decrease in activity at 37°C even in the presence of the optimal Brij concentration (Figure 3.10 c-d). Since the activity of both domains was maintained by 0.0015% Brij-35 at 22°C, all subsequent experiments were performed using these conditions.

The fluorescent signal of N- and C-domain was amplified approximately 10- and 16-fold, respectively, by the addition of Brij-35 (Figure 3.10 a-d). It thus appears that the surfactant monomers not only stabilize but also activate the truncated ACE domains.

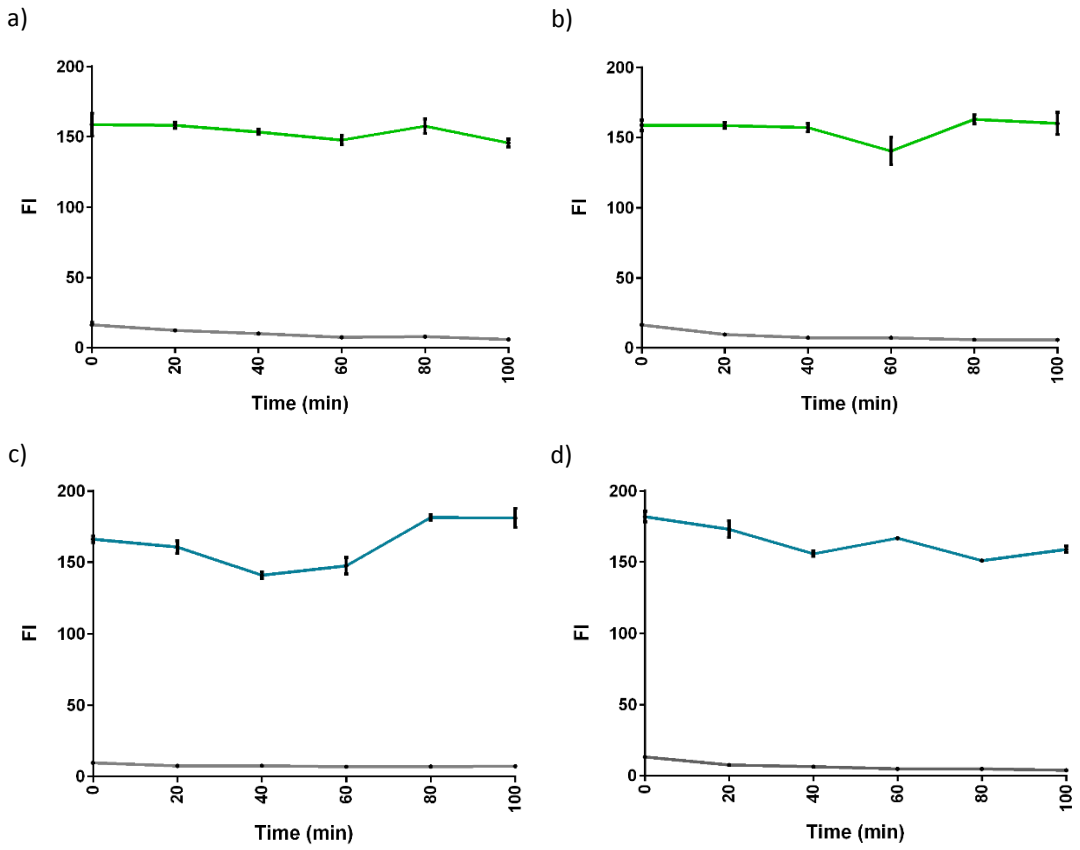


Figure 3.10: Enzyme stability as a function of time in the presence or absence of 0.0015% Brij-35.
a) and b) N-domain or c) and d) C-domain were incubated at 22°C (a) and c)) or 37°C (b) and d)) in the presence (blue or green, respectively) or absence (grey) of Brij-35. FI: fluorescence intensity. The time-course was performed inversely to enable simultaneous assessment of all samples' activity immediately after collection in a single assay. Data represented as mean ± SEM of triplicate wells.

3.4.5 Characterization of catalytic efficiency

The affinity of the two domains for the Hippuryl-His-Leu (HHL) substrate and the efficiency whereby they catalyze its hydrolysis was determined from initial reaction velocities (where substrate hydrolysis was less than 10%) as a function of substrate concentration. Fitting of the data to the non-linear Michaelis-Menten model revealed that substrate inhibition occurred at 2.4mM HHL (Figure 3.11). This was particularly pronounced for the C-domain. Exclusion of this data point resulted in a good fit (according to the R^2 fitting coefficients) of the initial velocities to the Michaelis-Menten model.

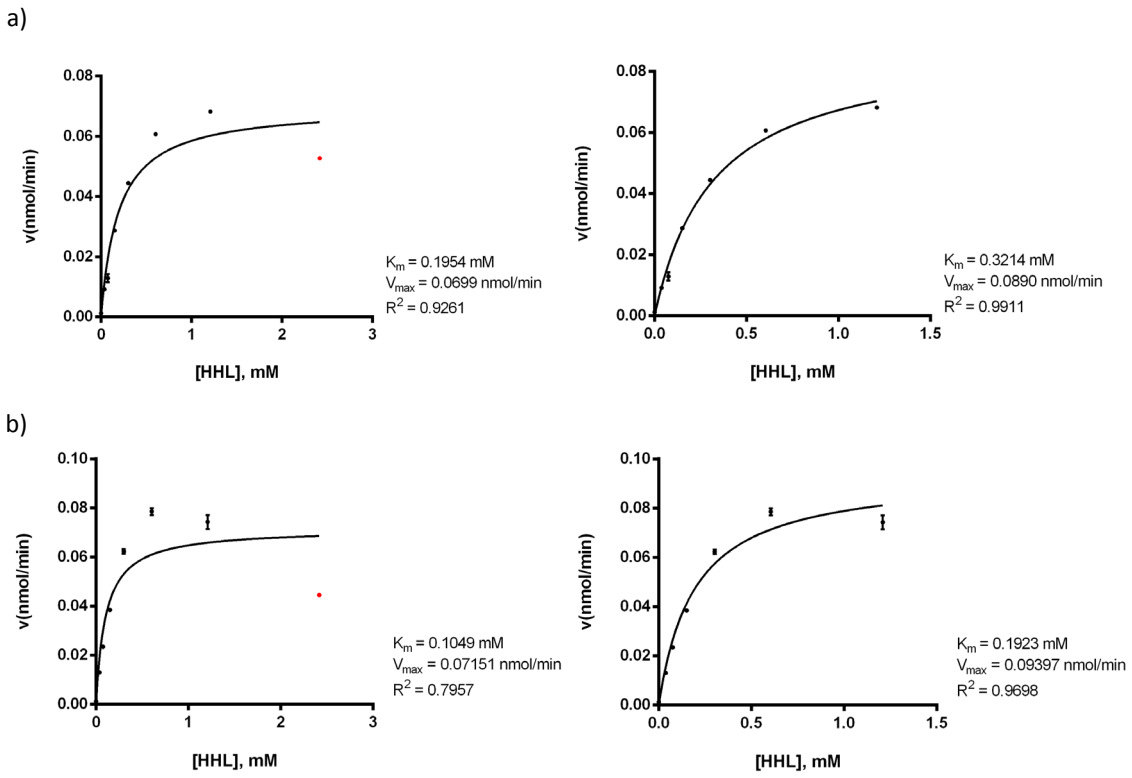


Figure 3.11: Michaelis-Menten kinetics for hydrolysis of HHL substrate.
Velocities are given for a) N-domain and b) C-domain with substrate inhibition at higher concentrations indicated as red data points. Michaelis-Menten parameters were determined using GraphPad Prism® v6.01 after exclusion of the last data point as shown on the right-hand panels. Data represented as mean ± SEM of triplicate wells.

The parameters describing HHL binding affinity (K_m), rate of hydrolysis (k_{cat}) and overall catalytic efficiency (k_{cat}/K_m) could thus be determined for the N- and C-domain (Table 3.1). The substrate binding affinity was non-domain selective whereas catalysis by the C-domain was slightly faster. Overall, the C-domain was approximately 4-fold more efficient than the N-domain at hydrolysing

HHL. The N-domain k_{cat} compared well with that reported in literature under conditions of potassium phosphate buffer pH 8.3, 300mM NaCl and 10µM Zn (Woodman *et al.*, 2006). Under the conditions used here, however, the C-domain k_{cat} was reduced and the binding affinity to both domains increased.

Table 3.1: Catalytic efficiency of HHL hydrolysis by N- and C-domain.
Data was analysed with the Michaelis-Menten model for determination of K_m and k_{cat} values (GraphPad Prism® v6.01) and represented as mean ± SEM of three independent experiments.

Construct	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}. mM^{-1}$)
N-domain	0.266±0.028	19.91±0.46	76.59±7.97
C-domain	0.277±0.011	84.34±7.28	303.13±13.84

3.4.6 Characterization of inhibitor binding

Binding of lisinopril, GGH-lisinopril (GGHLis) and Cu-GGH-lisinopril (CuGGHLis) was characterized by determining the initial velocity of HHL hydrolysis across a range of inhibitor concentrations. The CuGGHLis used in this section was prepared at a 1:1 molar ratio of metal to chelator. Since ACE is inhibited by $CuSO_4$ (Friedland and Silverstein, 1976), no free Cu^{2+} should be present when characterizing the inhibitor's binding affinity. This was confirmed by the lack of absorbance at 750-800nm (Appendix A13).

For the assumptions of the Michaelis-Menten model to be valid, initial reaction velocities where less than 10% substrate hydrolysis has occurred should be used in kinetic analyses. A low enzyme concentration of 7.5nM N-domain or 2nM C-domain was therefore chosen. The slow rate of catalysis, however, necessitated a substrate concentration above K_m (but below the substrate inhibition point) to ensure detectable fluorescent signal during the 15-minute assay in the presence of inhibitor. An HHL concentration of 1mM was therefore chosen to characterize inhibitor binding. Each inhibitor's IC_{50} value was determined from a sigmoidal dose-response curve (Figure 3.12).

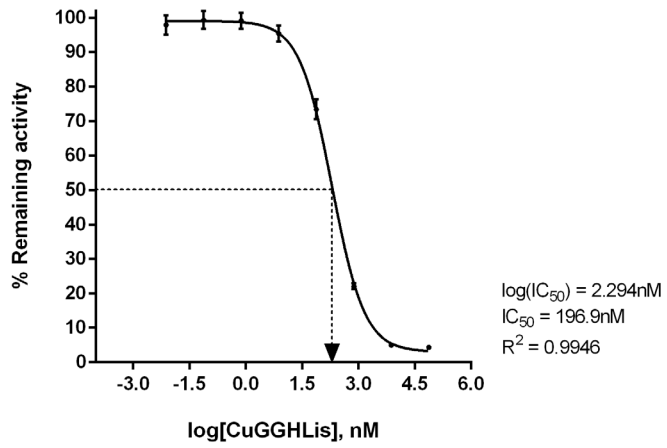


Figure 3.12: Representative inhibition curve for determination of inhibitor concentrations at which 50% enzyme inhibition occurs (IC_{50} values).
The N-domain was incubated with CuGGHLis and the percentage remaining activity fitted to a sigmoidal dose-response curve (represented as mean ± SEM of triplicate wells) to determine the IC_{50} value (arrow).

The inhibitors' binding constants (K_i values) were subsequently calculated using the Cheng-Prusoff equation where $K_i = IC_{50} / (1+[S]/K_m)$ and [S] = [HHL] = 1mM (Table 3.2). Kinetic characterization revealed low nanomolar inhibition of both domains by the three inhibitors tested. Lisinopril binding occurred with sub-nanomolar affinity and approximately 10-fold C-domain selectivity, presumably due to a strong amine interaction resulting from its more acidic prime subsite. GGHLis displayed decreased binding affinity for both domains and lacked domain selectivity. This likely resulted from an increase in inhibitor size and loss of the amine charge upon attachment of the GGH chelator. Following chelation of GGHLis to Cu^{2+} and formation of the bulky square planar metal-binding domain, a further decrease in binding affinity was observed. The slight C-domain selectivity suggests that the metallodrug is better accommodated by the larger C-domain prime subsite.

Table 3.2: Affinity of inhibitor binding to the N- and C-domain.
IC_{50} values were determined from sigmoidal dose-response curves and inhibitor binding constants (K_i values) calculated using the Cheng-Prusoff equation [a]. Data represented as mean ± SEM of three independent experiments.

	N-domain		C-domain		C-selectivity
Construct	**IC_{50} (nM)**	**K_i (nM)**	**IC_{50} (nM)**	**K_i (nM)**	**(K_i^{Ndom} / K_i^{Cdom})**
Lisinopril	4.09±0.16	0.86±0.03	0.40±0.02	0.09±0.004	9.6
GGHLis	30.57±3.36	6.41±0.71	46.03±3.76	9.99±0.82	0.6
CuGGHLis [b]	214.20±8.77	44.94±1.84	71.74±6.00	15.57±1.30	2.9

[a]: $K_i = IC_{50} / (1+[S]/K_m)$ where [S] = [HHL] = 1mM
[b]: metallodrug prepared at 1:1 chelator to metal ratio

The K_i values obtained for lisinopril corresponded well to those reported for the truncated domains of bovine ACE under similar conditions (Binevski *et al.*, 2003). GGHLis and CuGGHLis, however, displayed greater affinity to the truncated domains than reported by Hocharoen *et al* for full-length sACE using domain-selective substrates (Hocharoen *et al.*, 2013).

A higher concentration of inhibitor was likely required to observe complete inhibition of a particular domain in sACE form since the (Abz)-SDK(Dnp)P-OH and (Abz)-LFK(Dnp)P-OH substrates used were only 50- or 73-fold selective for the N- or C-domain, respectively (Rousseau *et al.*, 1995; Bersanetti *et al.*, 2004). Since both domains have the same affinity for substrate binding, a fraction of the observed percentage remaining activity of (Abz)-SDK(Dnp)P-OH, for example, could originate from the C-domain. This effect might have been exacerbated by the negative cooperativity that occurs upon inhibitor binding where binding of one molecule of inhibitor to sACE decreases the affinity of the second domain for inhibitor binding (Binevski *et al.*, 2003). Since these effects do not apply to the truncated domains, the IC_{50} and K_i values determined here likely provide a more accurate description of the inhibitors' binding affinities.

3.4.7 Determination of the mode of inhibition

Due to the disparities in inhibitor binding affinity between full-length and truncated ACE, it was investigated whether CuGGHLis competed with substrate for binding to the truncated single-domain proteins' active sites, as reported by Hocharoen *et al* for sACE (Hocharoen *et al.*, 2013). Inhibition was assessed over a range of inhibitor concentrations using the Dixon method (Dixon, 1953). As evidenced by the intersection of the Dixon plots in the upper left quadrant, CuGGHLis acted as a competitive inhibitor of both domains (Figure 3.13).

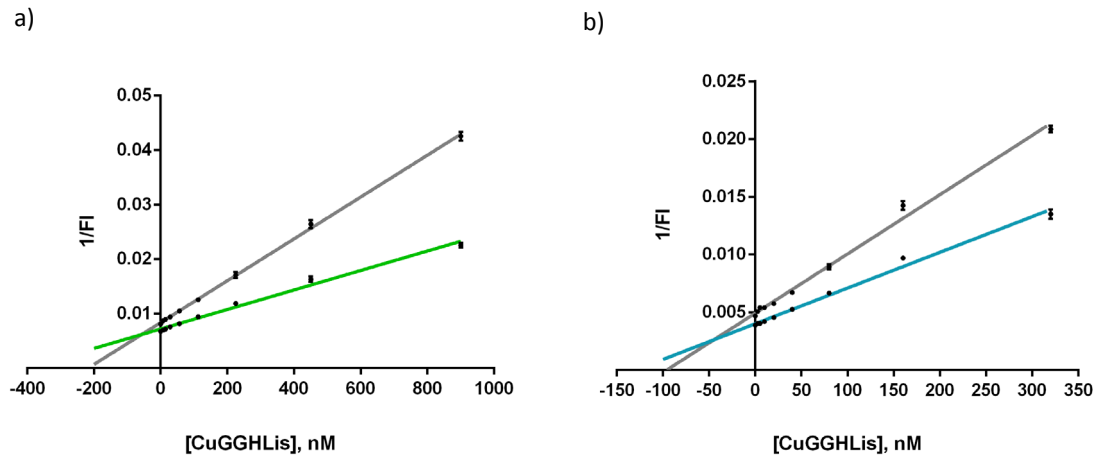


Figure 3.13: Dixon plots illustrating the mode of ACE inhibition by CuGGHLis.
The activity (FI) of a) N-domain or b) C-domain was assessed over a range of CuGGHLis concentrations using 1mM HHL (green or blue, respectively) or 0.5mM HHL (grey). Data represented as mean ± SEM of triplicate wells. FI: fluorescence intensity.

3.4.8 Time-dependent oxidative inactivation

The decrease in activity resulting from oxidative inactivation was monitored over 8 hours. Aliquots of samples containing protein in the absence or presence of CuGGHLis and/or redox co-reactants (H_2O_2 or ascorbate or H_2O_2 and ascorbate) were incubated at 22°C and the activity assessed at 20-minute intervals using HHL assays. The CuGGHLis used from this point onwards was prepared at a 1:1.1 molar ratio of metal to chelate and thus contained no free copper ions. The metallodrug was incubated with each domain at a concentration equal to its IC_{20} value (IC_{20} = IC_{50} x 2/5) so that a time-dependent decrease in activity resulting from oxidative catalytic inactivation could be observed from the initial approximately 80% enzyme activity at time zero. The CuGGHLis IC_{20} values for N- and C-domain were equal to 85.68nM and 28.7nM, respectively. All samples within a time-course were prepared from the same protein, inhibitor and co-reactant dilutions, incubated simultaneously and assayed in parallel to avoid introduction of technical variability. Activity at each time-point was expressed as a percentage relative to the control sample (enzyme containing no co-reactant or metallodrug) as a function of time.

Activity of the N- and C-domain samples containing no co-reactant or metallodrug was maintained for 8 hours (data not shown), thus confirming that 0.0015% Brij-35 successfully prevented thermal degradation at 22°C. The N-domain displayed no inactivation in the absence of CuGGHLis and presence of ascorbate or H_2O_2 (Figure 3.14 a). In the presence of both co-

reactants, however, an exponential decrease in activity to a plateau of 65% remaining activity was observed. Since the co-reactants were present in an approximately 10^6-fold excess over enzyme, the inactivation reaction was not limited by their concentration. Diffuse radicals thus did not completely inactivate the N-domain but only reduced its catalytic efficiency through oxidative modifications.

The C-domain was slightly inactivated by diffuse radicals in the presence of H_2O_2 but markedly so in the presence of both co-reactants (Figure 3.14 c). This diffuse-radical inactivation was faster than that of the N-domain (Figure 3.15) and ultimately led to a 57% loss of C-domain activity in the absence of CuGGHLis (Figure 3.14 c). The levels of activity remaining after diffuse-radical inactivation of the N- and C-domain (65% and 43%, respectively) were strikingly similar to those obtained by Michel *et al* (66% and 42%, respectively) after 4 hours' incubation with a 2,2'-azo-bis(2-amidino-propane) generator of hydroxyl radicals (1mM) (Michel *et al.*, 2001). This agreement with their study on the two domains' oxidation in full-length sACE form suggests that the data presented here on the truncated domains is physiologically relevant.

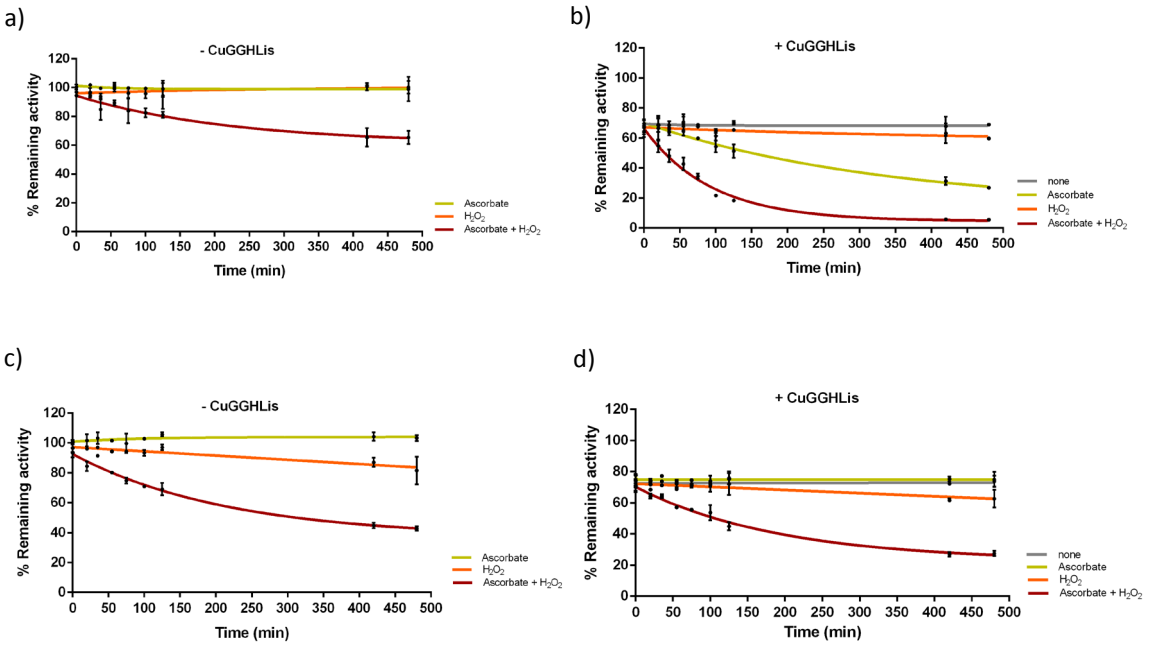


Figure 3.14: First-order exponential decay curves showing time-dependent catalytic inactivation. Remaining activity is given for N-domain (a-b) and C-domain (c-d) in the absence (a and c) or presence (b and d) of CuGGHLis at IC_{20}. Grey, orange, yellow and red curves correspond to samples containing no co-reactant, 1mM H_2O_2, 1mM ascorbate or 1mM H_2O_2 + 1mM ascorbate, respectively. Activity determined in triplicate wells at each time-point is represented as mean ± SD from two independent time-course experiments and expressed relative to enzyme only control.

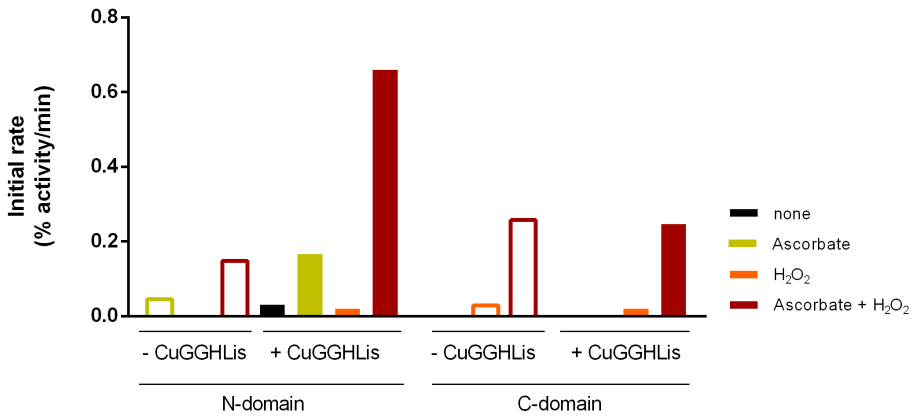


Figure 3.15: Initial rates of N- and C-domain inactivation in the presence or absence of CuGGHLis and redox co-reactants.
Yellow, orange and red correspond to samples containing ascorbate, H_2O_2 or ascorbate + H_2O_2. Initial rates were calculated from the average first-order exponential decay curves obtained from two independent time-course experiments.

Based on the observation that sACE was protected from oxidative inactivation by hydroxyl radical scavengers (dimethylthiourea or a combination of mannitol and methionine) or thiol-reducing agents (dithiothreitol or mercaptoethanol), it was proposed that hydroxyl radicals were the active ROS species (Michel *et al.*, 2001). The thiol-reducing agents' protection, however, did not offer much insight into the mechanism whereby diffuse radicals reduced ACE activity since the two active site cysteines are oxidized to a disulphide in the active enzyme (Sturrock *et al.*, 1996) and the single unbound cysteine resides on the protein surface. The results were further confounded by the fact that dithiothreitol addition, in itself, caused significant inhibition of ACE activity (Michel *et al.*, 2001).

Although over-oxidation of the disulphide by H_2O_2 or singlet oxygen oxidants into thiosulphinate, thiosulphonate, disulphide trioxide or disulphone species might occur (Kim *et al.*, 2015) (Figure 3.16), it is unclear whether such modifications would induce conformational change and ACE inactivation. Diffuse radicals might alternatively inactivate ACE through oxidation of the Zn-coordinating histidine residues. The greater susceptibility of the C-domain towards oxidation by diffuse radicals might be due to its more exposed active site (as discussed in Chapter 2) or its lower carbohydrate content compared to the N-domain. This will be discussed further in Chapter 4.

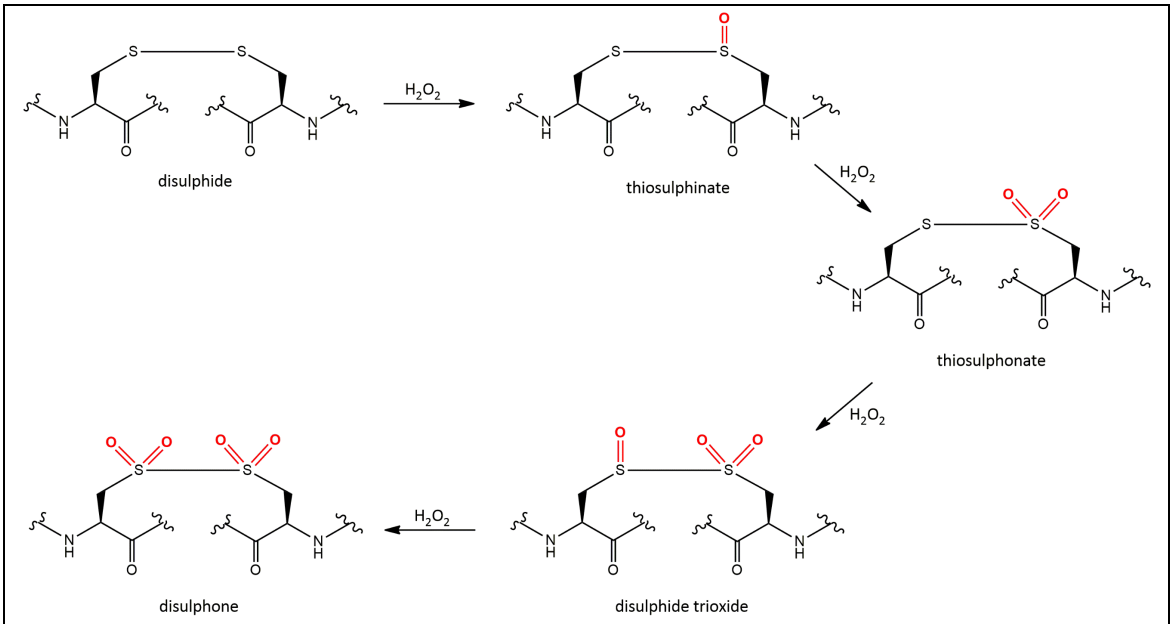


Figure 3.16: Mechanism of cysteine disulphide oxidation by oxidants such as hydrogen peroxide (H_2O_2) or singlet oxygen.
Adapted from Kim *et al* (Kim *et al.*, 2015).

It has been proposed that in the presence of a metallodrug and O_2, Fenton-type chemistry results in the formation and immediate reaction of metal-associated ROS at the site of production (Cowan, 2008). The ATCUN motifs, such as GGH, stabilize the $Cu^{3+/2+}$ redox couple (McDonald *et al.*, 1997) and are believed to cause enzyme inactivation via formation of an active $Cu^{3+}=O^{2-}$ oxone intermediate (Cowan, 2008). In the presence of CuGGHLis and H_2O_2 alone, a slight degree of N-domain inactivation was observed (Figure 3.14 b). Under these conditions, the Cu^{3+} state could be generated through a one-electron transfer reaction but since there is no efficient way of reducing the metal, redox cycling would not occur and the inactivation rate is very slow (Figure 3.15).

In the presence of CuGGHLis and ascorbate, N-domain inactivation was noticeably increased (Figure 3.14 b). Under these conditions, H_2O_2 would be formed from O_2 and the Cu^{3+} state generated by one-electron transfer (Figure 3.17 b). Following oxidation of the protein by the copper-associated ROS species, the Cu^{2+}-GGHLis state would be regenerated through one-electron transfer from ascorbate. The rate of inactivation is therefore increased over that observed in the presence of CuGGHLis and H_2O_2 alone through redox cycling (Figure 3.15). In the presence of metallodrug and both co-reactants, the most pronounced rate of catalytic inactivation was observed with only 18% N-domain activity remaining after 125 minutes'

incubation (Figure 3.14 b). Addition of H_2O_2 in the presence of ascorbate and O_2 would result in a mechanistic shunt to drive formation of the Cu^{3+} state, protein oxidation and subsequently redox cycling through reduction of the oxidized metal centre (Figure 3.17 b). This rapid copper-mediated oxidation ultimately resulted in complete N-domain inactivation. The initial inactivation rates determined from the exponential decay curves clearly illustrate that N-domain inactivation is enhanced over background (diffuse radicals in absence of inhibitor) in the presence of CuGGHLis (Figure 3.15).

In contrast, addition of CuGGHLis did not affect the rate of C-domain inactivation and 27% residual activity was retained, even after 8 hours' incubation with H_2O_2 and ascorbate (Figure 3.14 d). Previously, Joyner *et al* reported a similar plateau of approximately 30% remaining activity for the two-domain sACE after incubation with CuGGHLis (Joyner *et al.*, 2012). Thereafter, Hocharoen *et al* reported that the C- as well as the N-domain experienced CuGGHLis-mediated catalytic inactivation in the presence of both co-reactants, albeit at a reduced rate (18-fold greater N-domain initial inactivation rate) (Hocharoen *et al.*, 2013). The results presented here, however, indicate that the C-domain only undergoes oxidation by diffuse radicals and that CuGGHLis-mediated oxidation occurs rapidly and exclusively in the N-domain. This N-domain inactivation reaction is catalytically efficient as evidenced by the high second order rate constant (k_2) in the presence of ascorbate and H_2O_2 (Table 3.3). Since the N-domain was stable in 0.05% Brij-35 at 37°C, an oxidative inactivation time-course was also performed under these conditions to enable comparison with the second order rate constants determined by Joyner *et al* for full-length sACE using a non-domain selective substrate (Joyner *et al.*, 2012). At this temperature, the N-domain inactivation trends were the same but the k_2 values were increased (Appendix A16). The k_2 values determined under these conditions compared very well to those determined for the two-domain sACE, thereby providing further support for exclusive N-domain catalytic inactivation. This would therefore provide a highly selective therapy for the treatment of fibrosis at sub-stoichiometric inhibitor concentrations.

CuGGHLis is expected to be suitable for physiological use since previous studies showed that it induces production of metal-bound ROS as opposed to an abundance of diffusible ROS (Joyner *et al.*, 2012). The ROS species would therefore be located to the metallodrug's immediate environment in the active site and selective target inactivation would occur. This is further supported by the *in vivo* abundance of ATCUN motifs, such as GGH, used for copper and nickel transport (Lau *et al.*, 1974).

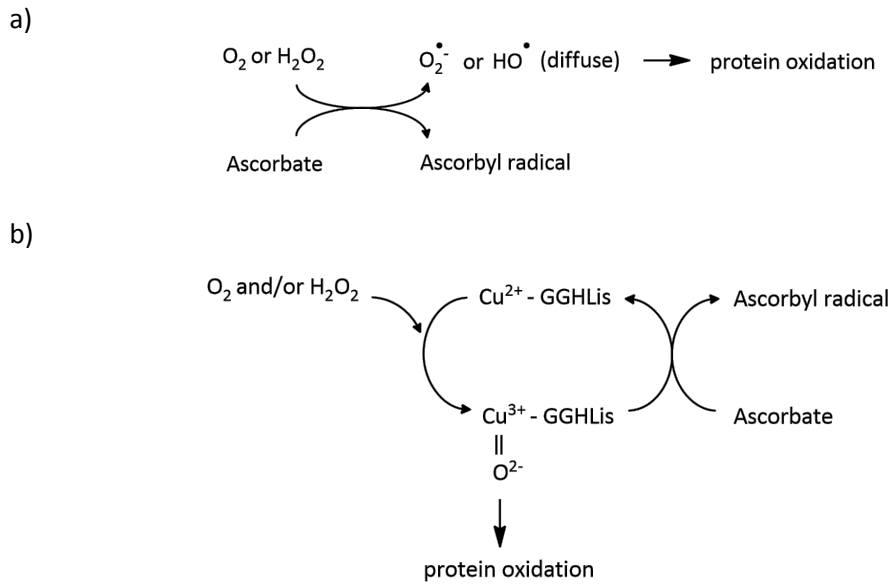


Figure 3.17: Mechanisms of radical generation in the presence of redox co-reactants.
a) Diffuse radical generation in metallodrug absence and b) metallodrug-mediated generation of metal-bound radicals in the active site.

Table 3.3: Initial rates and second order catalytic inactivation rate constants (k_2) determined at 22°C for N-domain in the presence or absence of CuGGHLis and redox co-reactants.
Initial rates were calculated from the average first-order exponential decay curves obtained from two independent time-course experiments performed in triplicate and rate constants as k_2 = (initial rate - background)/ [enzyme][CuGGHLis].

N-domain		Initial rate (nM min^{-1})		k_2 for inactivation (M^{-1} min^{-1})
		raw	- background	
- CuGGHLis	**Ascorbate**	0.0003		
	H_2O_2	-		
	Ascorbate + H_2O_2	0.0109		
+ CuGGHLis	**None**	0.002	0.002	3 625
	Ascorbate	0.012	0.009	14 128
	H_2O_2	0.001	0.001	2 310
	Ascorbate + H_2O_2	0.049	0.038	59 710

3.4.9 Kinetic characterization of residual activity

To investigate the mechanism of inactivation and basis for the differences between N- and C-domain, the residual activity was determined using a range of HHL concentrations after a 100-minute incubation at 22°C, in the presence or absence of redox co-reactants and CuGGHLis. The residual activity was subsequently characterized in terms of substrate binding affinity (K_m), turnover number (k_{cat}) and overall catalytic efficiency (k_{cat}/K_m). The 100min timepoint was chosen to have detectable inactivation but still enough residual activity for fluorescence signal detection and Michaelis-Menten kinetic characterization. All samples for both domains were prepared using the same inhibitor and co-reactant dilutions, incubated and assayed at the same time to avoid variability in substrate, inhibitor or co-reactant preparation between the samples or domains. All Michaelis-Menten curves were determined simultaneously for the 8 samples of each domain at the two time-points in triplicate wells of a 96-well plate.

One hundred minutes' incubation with co-reactants in the absence of metallodrug had no significant effect on the overall efficiency of N- or C-domain towards HHL hydrolysis (Figure 3.18). In the presence of CuGGHLis and ascorbate, however, a decrease in N-domain HHL binding affinity (increased K_m) and a slight decrease in turnover number (k_{cat}) was noticed (Figure 3.18 a). The N-domain catalytic parameters were not affected by H_2O_2 and CuGGHLis incubation but further addition of ascorbate led to a slight decrease in binding affinity and a significantly reduced turnover number. The N-domain's overall catalytic efficiency was thus significantly decreased after 100 minutes' incubation under these conditions. This suggests that CuGGHLis-mediated N-domain oxidative inactivation occurs either via sidechain oxidation of amino acids directly involved in HHL catalysis or oxidative cleavage of the protein backbone to ultimately reduce the number of moles of intact N-domain available for HHL hydrolysis. Oxidation of substrate-binding residues appears to play a less significant role.

No significant changes in C-domain catalytic parameters were detected after 100 minutes' incubation with co-reactants although the catalytic efficiency was slightly decreased in the presence of both co-reactants with or without CuGGHLis (Figure 3.18 b). Overall, this suggests that although CuGGHLis binds with similar affinity to both domains, the metal-chelate is optimally orientated in the N- but not the C-domain to catalyze oxidation of residues involved in substrate hydrolysis.

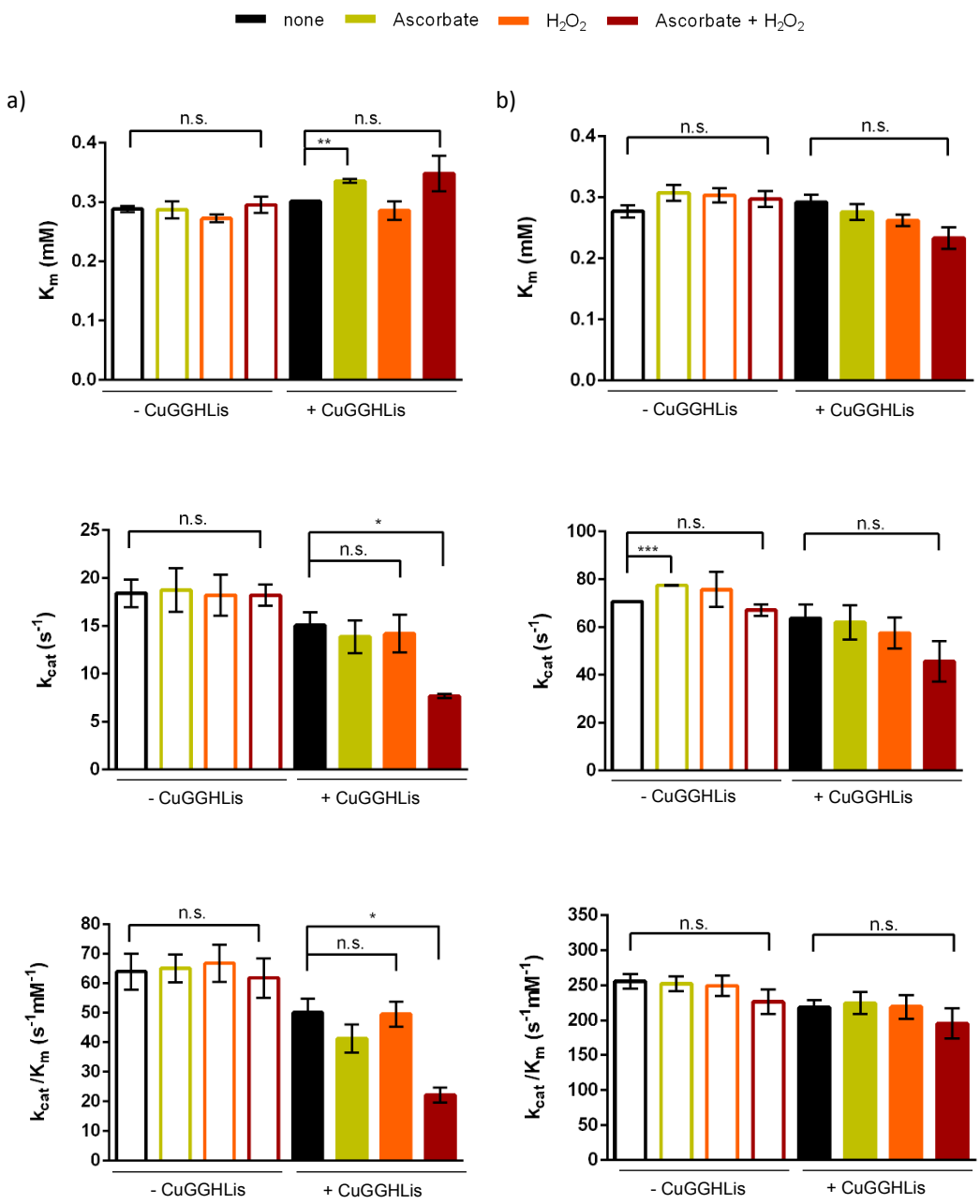


Figure 3.18: Characterization of residual activity of HHL hydrolysis after inactivation for 100 minutes. a) N-domain and b) C-domain. Kinetic parameters (K_m and k_{cat} values) determined in the absence and presence of CuGGHLis at IC_{20} are given as blank and filled bars, respectively. Black, yellow, orange and red bars correspond to samples containing no co-reactant, 1mM ascorbate, 1mM H_2O_2 or both (1mM ascorbate + 1mM H_2O_2), respectively. Parameters determined in triplicate wells across a range of HHL concentrations for each sample are represented as mean ± SD from two independent inactivation experiments. Analysed using an unpaired t-test (* $P < 0.05$); n.s: not significant.

Based on the molecular docking experiments performed with CuGGHLis (Joyner *et al.*, 2012) in the N- and C-domain, it appears that the metal may be near several readily oxidizable prime subsite residues (Figure 3.19). The metal-binding domain was orientated differently in the two domains and this might have occurred due to differences in the chemical properties or sidechain

lengths of unique S_2' residues (Figure 3.19 c-d). Molecular docking experiments performed with HHL (Douglas, 2011) further suggest that these oxidizable residues are involved in the hydrolysis of HHL since they are proximal to the C-terminal His-Leu (Figure 3.19 b). Flexibility in the CuGGHLis lysyl linker could allow the metal-chelate to transition slightly in the N-domain and cause oxidation of the critical oxyanion residues responsible for binding HHL's C-terminal carboxylate (Figure 3.19 c). In the C-domain, the metal-binding moiety is more distal from the oxyanion residues and therefore unlikely to affect HHL hydrolysis (Figure 3.19 d).

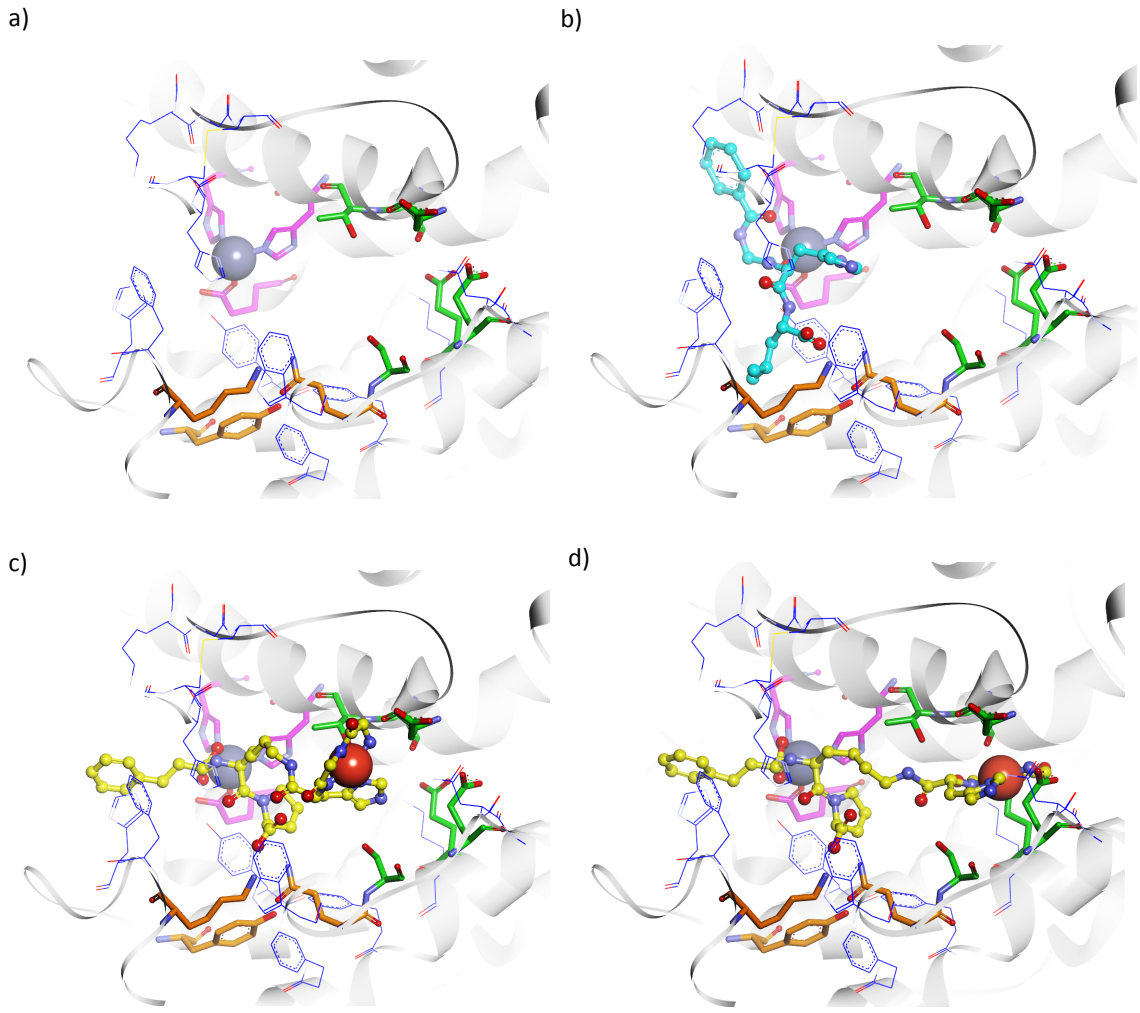


Figure 3.19: Structures of CuGGHLis and HHL docked into the N-domain.
Docking of HHL was performed by Dr R.G. Douglas (Douglas, 2011) and docking of CuGGHLis by Joyner *et al* (Joyner *et al.*, 2012). a) Oxidizable residues (blue lines) and unique residues in the N-domain prime subsite (green sticks); b) HHL (cyan) orientation in the N-domain, binding to the conserved oxyanion residues (orange sticks); c) CuGGHLis (yellow) orientation in the N-domain; and d) CuGGHLis (yellow) orientation in the C-domain. Zn is shown as a grey sphere, Cu as a brown sphere and Zn-binding residues in pink.

3.4.10 Assessment of artificial protease activity of CuGGHLis

To determine whether protein inactivation results from sidechain oxidation or protein backbone oxidation culminating in protein cleavage, SDS-PAGE was used. In this case protein was incubated for 7 hours at 22°C in the presence or absence of co-reactants and/or CuGGHLis to characterize the complete inactivation reaction. A sample which was not incubated was included to control for any possible thermal degradation products. No oxidative cleavage products were observed for either domain under any of the conditions tested (Figure 3.20). This suggests that sidechain oxidation is the main mechanism of irreversible ACE inactivation. In a two-domain protein such as sACE, this mechanism of action is desired since oxidative cleavage of one domain might indirectly affect the activity of the other due to their proximity.

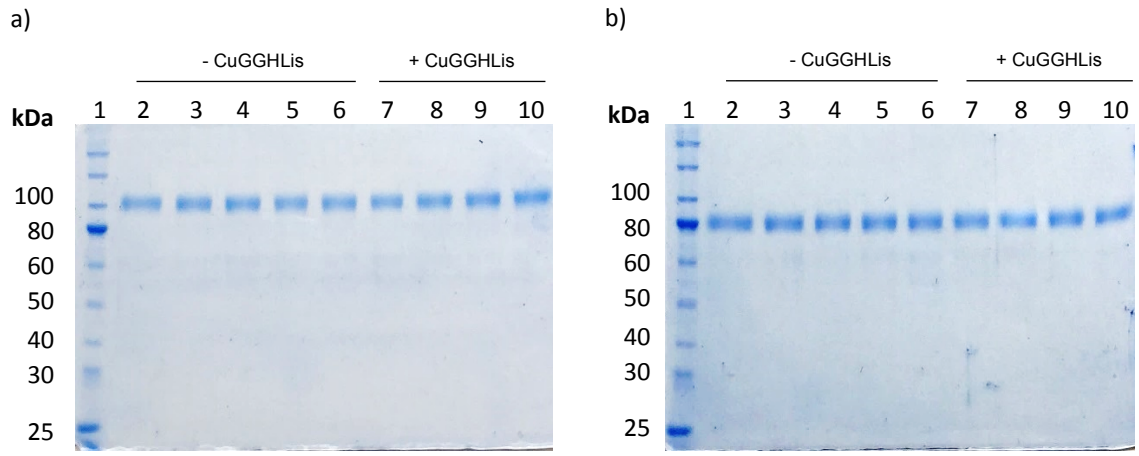


Figure 3.20: SDS-PAGE of inactivated ACE stained with colloidal Blue-Silver Coomassie.
a) N-domain or b) C-domain samples were incubated for zero minutes (lane 2) or 7 hours (lanes 3-10) in the presence or absence of CuGGHLis (at IC_{20}) and redox co-reactants at 1mM. Molecular weight markers (kDa) are given in lane 1. Lanes 2, 3 and 7: no co-reactant, lanes 4 and 8: ascorbate, lanes 5 and 9: H_2O_2, lanes 6 and 10: ascorbate + H_2O_2.

3.5 Conclusion

This study provides a careful analysis of the binding affinity of lisinopril, GGHLis and the CuGGHLis metallodrug by assessment of the two ACE domains in isolation with the same non-domain selective substrate. Upon incubation with a sub-stoichiometric amount of CuGGHLis in the presence of redox co-reactants rapid, complete and exclusive irreversible inactivation of the N-domain was observed. This analogue of the clinically approved ACE inhibitor lisinopril thus appears to be very promising for the treatment of fibrosis.

The basis for selectivity is likely a more optimal orientation of the CuGGH moiety to readily oxidizable residues in the N- than C- domain active site. Residual activity characterization and protein cleavage analysis suggests that inactivation mainly results from sidechain oxidation of N-domain residues involved in substrate hydrolysis. In future, the optimized conditions reported here could be used to explore the mechanism in more detail by utilizing X-ray crystallography and mass spectrometry to visualize the metal-chelate moiety's binding site and identify the oxidized residues' locations, respectively.

Chapter 4

Investigating the influence of *N*-glycans on ACE dynamics

4.1 Introduction

sACE is glycosylated at nine N-domain and six C-domain PNGs and the importance of these glycans in correct folding and processing of the enzyme has been shown experimentally (as discussed in Chapter 1 section 1.3.2). The molecular mechanism whereby they mediate folding, however, is unclear and has hampered efforts at obtaining a minimally glycosylated sACE glycoform amenable to crystallization. Elucidation of the sACE structure will offer insight into the mechanisms whereby negative cooperativity between the two domains and sACE dimerization occur upon inhibitor or substrate binding. Experimentally, it has also been shown that these glycans are important for ACE thermal stability (as discussed in Chapter 1 section 1.3.3). It is hypothesized that interactions of glycans at specific sites assist in folding by stabilizing flexible regions of the underlying protein and that this might contribute to thermal stability of the mature protein. Furthermore, it is proposed that the superior N-domain thermal stability could be the result of a decreased solvent accessible active site or protein surface area due to its higher carbohydrate content.

4.1.1 The role of glycosylation in oxidative inactivation

Decreased exposure of the protein surface to solvent could not only improve thermal stability but also protect against amino acid oxidative modifications by limiting ROS access (Li *et al.*, 2017). As described in Chapter 3, oxidative inactivation is the enzymatic activity loss observed upon oxidation of residues involved in substrate binding or hydrolysis. Although the active sites are ~90% identical (Corradi *et al.*, 2006), the C-domain is more susceptible than the N-domain to oxidative inactivation by diffuse radicals (Chapter 3 section 3.4.8) (Voronov *et al.*, 2002a; Michel *et al.*, 2001). Interestingly, the location of C-domain sites 2, 5 and 6 and N-domain sites 1, 2, 5, 7, 8 and 9 are unique to the respective domains (Figure 4.1). The 3D structure clearly shows that while the C-domain glycans (except site 6) are proximal to the N-terminal lid, the N-domain glycans are distributed around the molecule. A network of glycan-glycan interactions surrounding the N-domain could thus form a glycan-shield to prevent access of ROS to the active site.

In contrast, the sparser C-domain glycosylation and inter-subdomain repulsion (described in Chapter 2) could result in increased exposure of key active site residues to ROS.

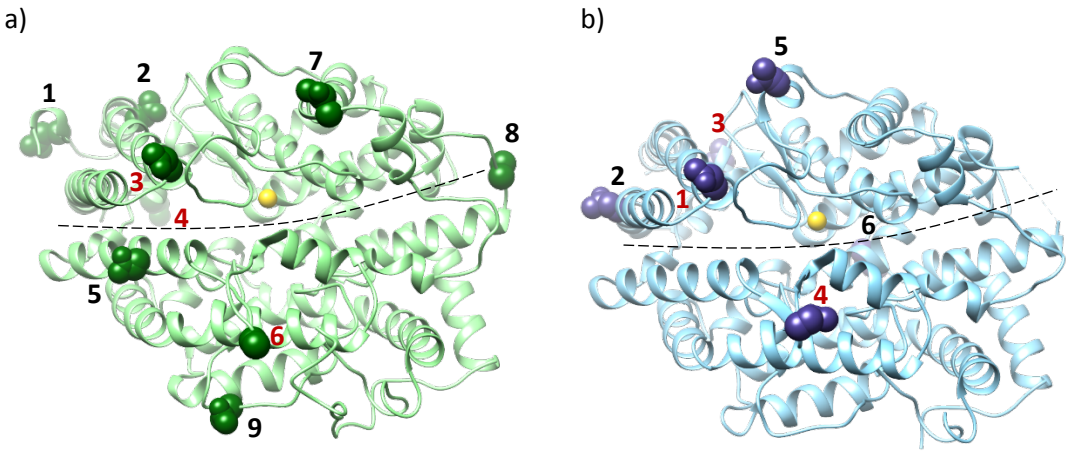


Figure 4.1: ACE crystal structures indicating PNGs.
a) N-domain (PDB ID: 2C6F) and b) C-domain (PDB ID: 1O8A) with PNGs indicated by spheres. Numbering of conserved sites is given in red. The active site cleft is indicated by the dotted line and the catalytic zinc ion as a yellow sphere.

Previously, Voronov *et al* reported that the C-domain (in sACE form) was greatly protected against water radiolysis-generated radicals by binding of the fluorescent hydrophobic dye 8-anilino-1-naphthalenesulfonic acid (1,8-ANS) (Voronov *et al.*, 2002b; Voronov *et al.*, 2003). The N-domain, which inherently is more resistant to ROS, was only minimally protected by 1,8-ANS. Although a conserved hydrophobic binding site was suggested from a modelled C-domain structure, the locations of these residues do not agree with the available crystallographic information. The binding site and mechanism of protection by 1,8-ANS is therefore unknown. Binding does appear to occur at a site distal from the active site since catalytic activity was unchanged. Interestingly, 1,8-ANS also did not affect thermal stability, suggesting that it does not bind to flexible loop regions. This molecule might function analogous to the unique N-domain glycan shield and protect C-domain residues involved in catalysis against ROS indirectly by facilitating active site cleft closure.

4.1.2 Implication of oxidative ACE inactivation for inflammation

Understanding the mechanism whereby 1,8-ANS offers protection against oxidation is of potential physiological relevance for the treatment of inflammation. In 1992 and 1993 Chen *et al* (Chen *et al.*, 1992; Chen and Catravas, 1993) observed that hydroxyl radicals generated

by phorbol 12-myristate 13-acetate (PMA)-activated neutrophils decreased ACE activity in cultured endothelial cells. Neutrophil incubation with 10mg/ml PMA (neutrophil: endothelial cell ratio of 1.25:1) stimulated superoxide anion release and completely inactivated ACE within 4 hours. In addition to ACE's role in regulating blood pressure via AngI cleavage, it also functions to cleave and inactivate bradykinin (BK), substance-P (SP), formyl-Met-Leu-Phe (fMLF) (Erdos, 1990) and N-acetyl-Ser-Asp-Lys-Pro (AcSDKP) (Rousseau *et al.*, 1995). In an inflammatory state ROS are produced by immune cells. Given the aforementioned ability of ROS to oxidatively inactivate ACE, it thus follows that ACE function would be impaired at the site of inflammation. Consequently, not only AngI but also BK, SP, fMLF and AcSDKP levels would be elevated (Figure 4.2).

Under basal conditions ACE is responsible for the majority of AngI cleavage and AngII production. An alternative pathway of AngII production via the serine protease mast cell chymase has been reported (Reilly *et al.*, 1982) but its contribution is minimal since chymase is stored in mast cell secretory granules as an inactive macromolecular complex. In an inflammatory state, however, mast cell degranulation leads to chymase activation, production of AngII, pro-inflammatory cytokines and ultimately a significant increase in leukocyte adhesion to endothelial cells (Company *et al.*, 2011). Similarly, neutrophil cathepsin G also has the ability to cleave AngI into AngII (Reilly *et al.*, 1982). Although oxidative inactivation of ACE prevents cleavage of AngI, this therefore does not decrease AngII levels. Increased AngII results in vasoconstriction, nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activation, production of superoxide anion, tumour necrosis factor alpha (TNF-α), interleukin 1 (IL-1), interleukin 6 (IL-6), interferon gamma (IFN-γ), transforming growth factor beta 1 (TGF-β_1) and increased vascular permeability via AT_1R binding (de Gasparo *et al.*, 2000). Although AngII binding to the angiotensin receptor type 2 (AT_2R) has an opposing effect (Siragy, 2000), this receptor is only expressed during development (Grady *et al.*, 1991). ACE2 cleavage of AngII also results in increased production of Ang1-7 with subsequent NO production and vasodilation via the MAS receptor (Simoes e Silva *et al.*, 2013). NO is anti-inflammatory under basal conditions but becomes pro-inflammatory when produced in high concentrations (Sharma *et al.*, 2007). The essentially irreversible reaction of NO with neutrophil-generated superoxide anion leads to peroxynitrite ($ONOO^-$) formation (Beckman and Koppenol, 1996). Interaction of proteins with this strong oxidant can result in tryptophan and cysteine oxidation, tyrosine nitration,

dityrosine formation and protein backbone oxidative cleavage (Ischiropoulos and al-Mehdi, 1995).

Vasodilation and NO production via the phospholipase C – protein kinase C – mitogen-activated protein kinase (PLC-PKC-MAPK) pathway is also mediated by binding of the BK nonapeptide to its constitutively and ubiquitously expressed bradykinin receptor B2 (B2R) (Scharfstein *et al.*, 2017). Furthermore, BK binding to B2R leads to increased tissue plasminogen activator (t-PA) levels (Brown *et al.*, 2000). Binding of t-PA to the low density lipoprotein receptor-related protein 1 (LRP-1) and Annexin-A2 receptors, not only increases M1 macrophage motility (Lin and Hu, 2014) and survival (Lin *et al.*, 2015) but also activates NF-κB, promoting M2 to M1 macrophage phenotypic change and inflammatory cytokine production (Lin and Hu, 2017). The effect of increased t-PA might be off-set, however, by a chymase AngII-mediated increase in plasminogen activator inhibitor type 1 (PAI-1) expression (Fogari and Zoppi, 2006). Nevertheless, increased PAI-1 is not beneficial since inactivation of t-PA along with TGF-β_1 stimulation by AngII leads to tissue fibrosis via fibrin accumulation, migration and adhesion of inflammatory cells and fibroblasts (Ghosh and Vaughan, 2012).

The bradykinin receptor B1 (B1R), on the other hand, is present at low levels in healthy tissues but rapidly induced upon tissue injury (Scharfstein *et al.*, 2017). Following BK cleavage by glycosylphosphatidylinositol-linked carboxypeptidase M (CPM), Des-Arg^9-BK binds to B1R and increases neutrophil migration to the site of inflammation, leukocyte adhesion to endothelial cells (Figueroa *et al.*, 2015), production of pro-inflammatory cytokines and hyperalgesia (Ahluwalia and Perretti, 1999). BK also induces mast cell degranulation which, in addition to the aforementioned AngII-mediated pro-inflammatory effects, leads to increased vascular permeability and release of prostaglandin D_2 (PGD_2), leukotrienes, platelet-activating factor (PAF), SP, NO and histamine (Theoharides *et al.*, 2012).

Increased SP similarly induces vasodilation, mast cell degranulation, cytokine release, neutrophil recruitment and activation (Katsanos *et al.*, 2008), angiogenesis (Kohara *et al.*, 2010) and t-PA production (Labinjoh *et al.*, 2001). Hyperalgesia in this case results from SP binding to the Neurokinin-1 receptor (NK-1), PLC-PKC signalling and ROS production (Linley *et al.*, 2012).

The f-MLF peptide plays a role in inflammation by increasing neutrophil and macrophage recruitment upon binding to its formyl peptide receptor (FPR) or FPR-like 1 receptor on the surface of these cells (Schiffmann *et al.*, 1975). Increased adherence to endothelial cells, macrophage activation via PKC and protein tyrosine kinase (PTK) (Shrivastava, 2007) and subsequent pro-inflammatory cytokine, superoxide anion, H_2O_2 and NO production also result.

In contrast to the pro-inflammatory effects of AngII, BK, SP and fMLF, increased AcSDKP levels upon oxidative ACE inactivation decreases inflammation and fibrosis. AcSDKP reduces culture bone marrow stem cell maturation to mature macrophages, macrophage migration, TNF-α secretion and collagen deposition (Sharma *et al.*, 2008). This effect, however, might be overshadowed by the overwhelmingly pro-inflammatory events since it would still be selectively cleaved into inactive peptides by the less ROS-susceptible N-domain (Rousseau *et al.*, 1995). The ROS-susceptible C-domain has previously been shown to hydrolyse AngI, BK and SP more efficiently than the N-domain (Wei, 1991; Jaspard *et al.*, 1993).

Therefore, it is likely that swelling, angiogenesis, cell proliferation, cytokine production, neutrophil recruitment, macrophage activation and an increase in pain perception occurs by the collective action of AngII, BK, SP and fMLF. Previously, a dose-dependent relationship was described between the degree of ACE inactivation and the number of immune cells, their activation and adherence to endothelial cells (Chen *et al.*, 1992). The pro-inflammatory effects of these peptides would thus initiate a continuous cycle of ROS production by immune cells and ACE inactivation, thereby promoting the progression of inflammation and development of fibrosis (Figure 4.2).

Understanding the mechanism of C-domain protection by 1,8-ANS could facilitate design of an anti-inflammatory 'ACE-protecting agent' for co-administration with antioxidants and an N-selective ACE inhibitor in inflammatory disease. Since binding supposedly occurs distal from the active site, no alteration in blood pressure or basal ACE expression is anticipated. Although the ROS-protected C-domain active site would be free to cleave AngI and produce AngII, in a normotensive state with basal ACE expression the cleavage of BK, SP and fMLF into inactive peptides and build-up of AcSDKP would overall have an anti-inflammatory outcome, interrupting the cycle of ROS production and ACE inactivation.

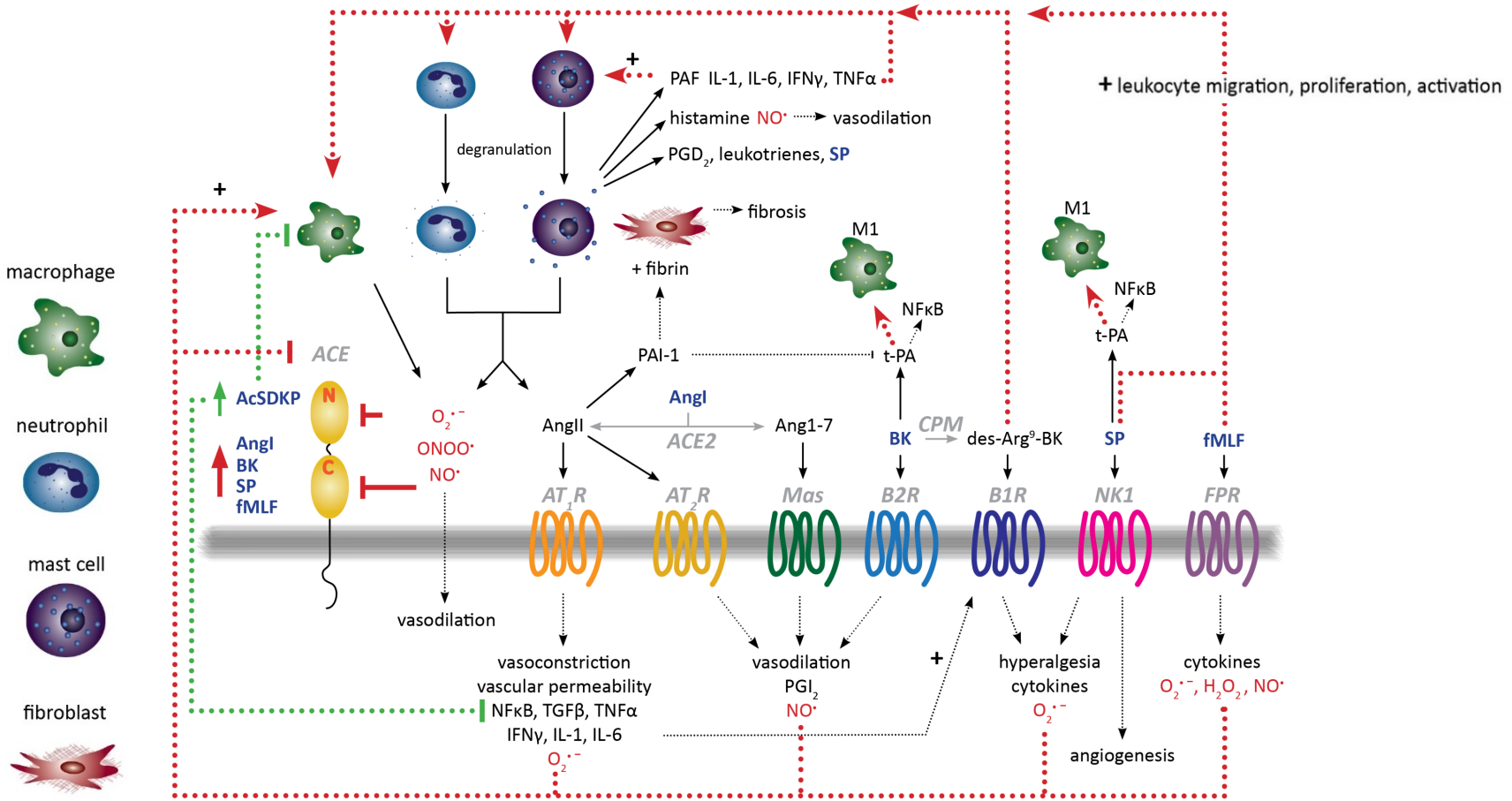


Figure 4.2: Proposed cycle at the site of inflammation whereby ACE oxidative inactivation leads to fibrosis and progression of inflammation.
Pro- and anti-inflammatory processes are indicated by red and green dotted lines, respectively. AngI, BK, SP and fMLF refer to ACE substrates that bind to their respective receptors (labelled grey). ACE: angiotensin converting enzyme; ACE2: angiotensin converting enzyme 2; AT_1R: angiotensin receptor type I; AT_2R: angiotensin receptor type II; B1R: bradykinin receptor B1; B2R: bradykinin receptor B2; NK1: neurokinin-1 receptor; FPR: formyl peptide receptor.

4.2 Aims and objectives:

This part of the study firstly aimed to determine the molecular mechanism whereby ACE N- and C-domain glycosylation confers thermal stability. Secondly, the C-domain's increased susceptibility to oxidative inactivation and protection by 1,8-ANS was investigated. Insight gained by achieving these aims would in future assist sACE structure determination and design of novel ACE-modulating agents with potential anti-inflammatory activity.

Molecular dynamics simulations on the fully glycosylated truncated N- and C-domains were employed to address the following objectives:

1. Determine the mechanism for site-specific glycosylation in thermal stabilization by analysing protein flexibility, identifying glycan-protein interactions and evaluating the volume occupied per glycan.

2. Investigate the molecular basis for the N-domain's superior resistance to oxidative inactivation by identifying glycan-glycan interactions and analysing active site solvent access.

3. Elucidate the mechanism whereby 1,8-ANS binding to the C-domain protects against oxidative inactivation by identifying its binding site.

4.3 Methodology:

4.3.1 Crystal structure selection

Glycan attachment in eukaryotic cells occurs at a potential *N*-linked glycosylation site (PNGs) with sequence N-X-T/S (where X is any amino acid except proline) which remains solvent-exposed after protein folding. Biologically relevant glycosylation sites are therefore solvent exposed. In addition to experimental evidence, *in silico* assessment of potential *N*-linked sites' solvent accessibility can therefore guide the decision of which sites should be glycosylated. High-resolution crystal structures of the N- and C-domain of ACE (in minimally glycosylated forms to facilitate crystallization) have previously been solved. In the current study, however, these structures could not be used since N25Q mutation used in constructing the minimally glycosylated N-domain form (Anthony *et al.*, 2010) caused reorientation and burying of the side-chain. The lower resolution apo structures solved from fully glycosylated protein expressed in the presence of NBDNJ were therefore used instead (Corradi *et al.*, 2006; Natesh *et al.*, 2003) (N-domain PDB ID:2C6F, 3.01Å resolution; C-domain PDB ID: 1O8A, 2.00Å resolution). The single C-domain molecule and molecule A of the N-domain asymmetric unit were chosen.

4.3.2 Protein structure preparation

The lacking C-domain N-terminal L^{37}-V-T^{39} and C-terminal Y^{619}-NWTPN-S^{625} residues (standard tACE numbering (Ehlers *et al.*, 1989)) were built using Modeller (Sali and Blundell, 1993) in UCSF Chimera v1.10.1 (Pettersen *et al.*, 2004) (Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311)). Since Modeller renumbered the residues during this process (for example L^{37}-V^{38}-T^{39} to L^{1}-V^{2}-T^{3}), all references to C-domain from this point onwards are 36 less than the standard tACE residue numbering. The glycan site numbering used in this Chapter is given in Table 4.1.

Table 4.1: C-domain glycan numbering used in the glycosylated MD structures.

		Standard tACE numbering	Numbering used in this chapter
N-terminal residue	Leu	37	1
Glycan site	c1	N72	N36
	c2	N90	N54
	c3	N109	N73
	c4	N155	N119
	c5	N337	N301
	c6	N586	N550

Standard tACE numbering is according to the testis ACE isoform which includes a unique 36-residue N-terminal region (absent in sACE). The constructs used for X-ray crystallography lacked this region and thus the first N-terminal amino acid was Leu37. In this chapter, Modeller renumbered this construct to start at 1, thus numbering used is 36 less than standard tACE numbering.

All subsequent N- and C-domain methodologies for protein preparation, simulation and analysis were analogous. Sugar, polyethylene glycol, N-carboxyalanine and acetate molecules were deleted prior to protein pre-processing with the Schrödinger Suite's (version 2014) Maestro Protein Preparation Wizard (Madhavi Sastry *et al.*, 2013). Bond orders were assigned, hydrogen atoms added for pH 7.0, disulphide bonds created and missing sidechains filled in with Prime. The Interactive hydrogen bond optimizer was used to visually inspect and adjust the histidine, asparagine (except *N*-linked) and glutamine tautomers for optimal hydrogen bonding. All crystallographic waters, except for the N-domain's Zn-coordinated water, were deleted after successful protein optimization.

4.3.3 Protein glycosylation and optimization

The Glycoprotein Builder contained within the GLYCAM Web-tools suite (Woods Group (2005-2018) GLYCAM Web; Complex Carbohydrate Research Centre, University of Georgia, Athens, GA. (http://glycam.org)) was used to build and covalently link the 3D structure of a bisialylated fucosylated biantennary complex type glycan (Figure 4.3) (DNeup5Acα2-3DGalpβ1-4DGlcpNAcβ1-2DManpα1-6[DNeup5Acα2-3DGalpβ1-4DGlcpNAcβ1-2DManpα1-3]DManpβ1-4DGlcpNAcβ1-4[LFucpα1-6]DGlcpNAcβ1-OH) to *N*-linked ACE glycosylation sites. This glycan structure was chosen since it was supported by previous mass spectrometric analysis of the truncated C-domain expressed in CHO-K1 cells (Yu *et al.*, 1997)

and quantitative HPLC analysis on human kidney sACE and tACE expressed in CHO-K1 cells (Ehlers *et al.*, 1992).

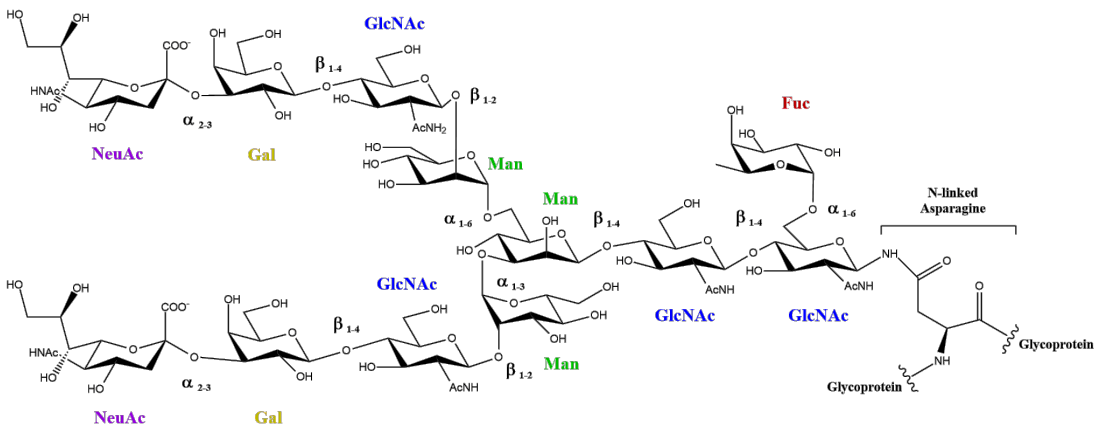


Figure 4.3: Stereochemical structure of the glycan attached to N- and C-domain PNGs. Glycosidic linkage types are indicated by Greek symbols and the monosaccharide names given in standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015).

The chloride, zinc and active site water were removed during the process of glycosylating the nine N-domain and six C-domain PNGs. GLYCAM automatically performs 1000 steps of steepest descent minimization on the attached glycans while keeping the protein atoms restrained. An error in the script (personal communication with GLYCAM developers), however, caused minimization of protein atoms in the absence of metal coordination. The Zn-coordinating residues' sidechain orientations in the resulting glycoprotein structure were thus altered and unsuitable for Zn binding. After glycosylation the coordinates for H361/347, H365/351, E389/375 (N-domain/C-domain numbering), Cl and Zn were therefore manually replaced by those from the protein structures optimized for hydrogen bonding. Additionally, a Zn-coordinating water molecule was built into both domains from its N-domain crystallographic coordinates.

UCSF Chimera v1.10.1 (Pettersen *et al.*, 2004) was used to detect steric glycan-glycan and glycan-protein clashes. These were manually resolved by altering the Asn-GlcNAc glycosidic linkages' phi (φ: O_5-C_1-N_{D2}-C_G), psi (ψ: C_1-N_{D2}-C_G-C_B) or chi (χ_1: N-C_A-C_B-C_G; χ_2:C_A-C_B-C_G-N_{D2}) torsion angles, as previously described (Li *et al.*, 2017; Wood *et al.*, 2013). Importantly, these

Asn-GlcNAc linkage angles were kept within allowed values (Petrescu *et al.*, 2004) to ensure biologically relevant presentation of the sugar to the protein.

To differentiate the *N*-linked asparagines, their residue names were changed to NLN in all relevant files. Similarly, the metal site residues were renamed HIR, HIT, GLR, WAR and ZN1 for H361/347, H365/351, E389/375 (N-domain/C-domain numbering), active site water and Zn, respectively.

4.3.4 Preparation of coordinate and topology files

All simulations were performed using AMBER 14 and AmberTools 15. The glycosylated proteins' coordinate and topology files were generated using the *tLEaP* module with AMBER ff14SB forcefield (Maier *et al.*, 2015) for the protein and GLYCAM_06j-1 forcefield (Kirschner *et al.*, 2008) for the glycan atoms. Bond, angle and charge parameters for the Zn-site (residue library and forcefield parameter modification files) were obtained from Dr N.F. Brás (University of Porto) (Brás *et al.*, 2014). The custom residues of the metal site were differentiated from those present in the AMBER ff14SB forcefield by renaming them as follows: first His = HIR, second His = HIT, Glu = GLR, Zn-coordinating water = WAR. The atom names were also changed accordingly (NK = HIR NE2; NN = HIT NE2; ON = GLR OE1; OK = WAR O; Z1 = ZN1). Neutralization was achieved by addition of 33 and 26 Na^+ ions to the N- and C-domain, respectively. The glycoproteins were subsequently explicitly solvated by immersion in a pre-equilibrated isometric truncated octahedron TIP3P water box (Price and Brooks, 2004) with a 10Å barrier between the outer atoms and the box boundary (58504 TIP3P or 30591 TIP3P to yield volumes of 2056181 $Å^3$ and 1155503 $Å^3$ for N- and C-domain, respectively). The final neutralized and solvated systems consisted of 188042 N-domain and 103072 C-domain atoms. Glycan, disulphide and metal bonding information was manually added to the *tLEaP* input script. The bonded model, previously reported by Brás *et al* (Brás *et al.*, 2014), was initially employed to describe Zn-coordination. With the conditions used here, however, this resulted in Van der Waals overlap between the WAR OK and the GLR ON atoms during minimization. A hybrid bonded/non-bonded model was therefore adopted where HIR, HIT and GLR were explicitly bound to ZN1, but WAR non-bonded and harmonically restrained. This is similar to that used in the recently parameterized M1 ZAFF (Yang *et al.*, 2017b). The lower- and upper-bound linear response regions were set to 0.056Å and 4.056Å, respectively, with an equilibrium bond length of 2.056Å and force constant of 66.54kcal/mol (Brás *et al.*, 2014). *Cpptraj* was used to centre the protein prior to simulation.

4.3.5 Glycoprotein molecular dynamics simulations

The following procedure was followed using the *sander* module to allow equilibration of the systems' energy, temperature and pressure: (a) 2500 steps steepest descent and 2500 steps conjugate gradient solvent minimization while imposing a 50.0kcal/mol restraint on the solute atoms; (b) 10000 steps steepest descent and 10000 steps conjugate gradient unrestrained minimization; (c) constant volume solvent heating from 10.0K to 300.0K in three stages over 600ps with the Langevin thermostat, while imposing a 50.0kcal/mol restraint on the solute atoms; (d) 200ps solvent equilibration at 300K and 1bar pressure using the Berendsen barostat with isotropic position scaling and a pressure relaxation time of 2.0ps, while imposing a 10.0kcal/mol restraint on the solute atoms; and (e) 200ps unrestrained system equilibration at 300K and 1bar pressure using the Berendsen barostat with isotropic position scaling and a pressure relaxation time of 2.0ps. Finally, unrestrained production dynamics were performed using the Particle Mesh Ewald Molecular Dynamics (PMEMD) code with GPU-acceleration. A total of 3000 trajectory frames taken at 10ps intervals and including all atoms were collected for each glycoprotein during 30ns of post-equilibration dynamics under physiological conditions (1bar, 300K).

Long-range electrostatic interactions were calculated using the Particle Mesh Ewald method with a non-bond interaction cut-off distance of 10.0Å. A 2fs timestep was used throughout with the SHAKE restraint applied during heating, equilibration and production dynamics. The wall-clock time was used as seed for the pseudo-random number generator during dynamics simulations.

4.3.6 Trajectory analysis procedure

Unless otherwise noted, all analyses were performed using *cpptraj* in AmberTools15. The systems' energies, volume, density and pressure were monitored using the *process_minout.perl* and *process_mdout.perl* scripts of *cpptraj*. Trajectory frames where these properties were equilibrated were subsequently analysed after stripping of TIP3P and Na^+ atoms. Root mean square deviation (RMSD) of all non-hydrogen or protein backbone atoms was assessed and time-averaged structures created by mass-weighted fitting against the first equilibrated frame. The root mean square fluctuation (RMSF) of backbone protein atoms was also assessed.

To ensure biological relevance of the obtained glycan conformations and evaluate sampling of conformational space, variations in *α1-6*, *α2-3* and Asn-GlcNAc glycosidic linkage torsion angles were monitored over time. The glycans' behaviour was evaluated by extracting long-range correlated motions of all non-hydrogen atoms using principle component analysis (PCA). Direction and magnitude of motion was visualized as porcupine plots and pseudo-trajectories of the first 3 modes with the Normal Mode Wizard in VMD v.1.9.3 (Humphrey *et al.*, 1996). Furthermore, VMD v.1.9.3 (Humphrey *et al.*, 1996) was used to assess how these motions contribute to protein shielding by calculating glycan occupancy VolMaps. After aligning the protein backbone CA atoms of all frames to the first equilibrated frame, occupancy VolMaps were computed for each glycan. A three-dimensional grid encompassing the glycan heavy atoms was constructed with a voxel (grid-point) size of 1Å. The atoms were treated as spheres based on atomic radii and assigned a status of 'occupied' if its atomic sphere overlapped with a voxel. Fractional occupancy of each voxel was calculated by normalizing occupancy over all frames to 1, written to volume maps and visualized as isosurfaces. To evaluate the dynamic range of glycan motion, the volume occupied per glycan was quantified using the Volutil package to convert fractional maps to binary volume maps. Values of '0' and '1' were assigned to each unoccupied and occupied voxel, respectively. Occupied volume was then quantified as the sum of all occupied voxels per glycan.

To describe the probability of possible glycan-glycan interactions identified from occupancy VolMaps, radial distribution functions were calculated for each glycan pair. The sugar residues involved in these contacts were subsequently identified by hydrogen bond analysis. Similarly, interactions between each glycan and the protein were identified and their frequency evaluated by hydrogen bond analysis over all frames. Clustering of the glycan conformations was used to extract the most representative structure for visualization of these hydrogen bonds. Five clusters were created for glycan heavy atoms of each domain using a hierarchical agglomerative approach with an epsilon of 15.0 and the average-linkage algorithm.

Differences between the domains' active site solvent accessibility were visualized by computing solvent tunnels using Caver Analyst v1.0 (Kozlikova *et al.*, 2014). Owing to limitations in computational resources, tunnel dynamics were only calculated over the last

100 frames (30th ns). The active site Zn was set as the starting point for tunnel detection with a 0.9Å minimum probe radius, 3Å shell radius and 4Å shell depth. Glycan atoms were included and WAR, ZN1 and Cl^- excluded from tunnel computation with all remaining settings kept to default. Resulting tunnels were visualized in PyMOL v2.0.6 (Schrodinger, 2017) and VMD v.1.9.3 (Humphrey *et al.*, 1996).

The Caver Analyst calculation is useful for identifying gating residues but does not, however, consider the tunnels' chemical properties. The newly developed AQUA-DUCT (Magdziarz *et al.*, 2017) tool was therefore used to quantify tunnel usage during the entire 30ns simulation of explicitly solvated N- and C-domain. The full trajectory was screened for any TIP3P water molecule that entered or exited an object region via the scope region. The object was defined as water molecules within a 3Å spherical zone around the centre of masses of the Zn-binding histidine, glutamate and water residues as well as a facing tyrosine (Y501 in N-domain and Y487 in C-domain). This was contained within the scope defined as the interior of a convex hull of C_α atoms. The identified molecules were then traced throughout the trajectory, their paths calculated and inlets at the scope boundary identified. Paths were trimmed to the protein surface using the Auto Barber procedure with a maximal cutting sphere radius of 4.0Å. Inlets were clustered to identify tunnels using the Barber method with automatic detection of outliers (small clusters of 2 inlets). Average paths were calculated and smoothed along with raw paths using a combined Distance Window and Max Step method (dwin_mss) with 5.0Å window and step sizes for convenient visualization in PyMOL v2.0.6 (Schrodinger, 2017).

4.4 Results and Discussion

4.4.1 Protein glycosylation and optimization

Amino acid numbering used in this chapter is explained in section 4.3.2 and Table 4.1. Following attachment of the 3D glycan structures to PNGs in GLYCAM, protein-glycan and glycan-glycan clashes were resolved by manually adjusting the torsion angles shown in Figure 4.4. The torsion angles of the final structures are given in Table 4.2 and corresponded well to literature values for this linkage (Petrescu *et al.*, 2004). Presentation of the core saccharide to the protein in the MD simulation input structure was thus biologically relevant.

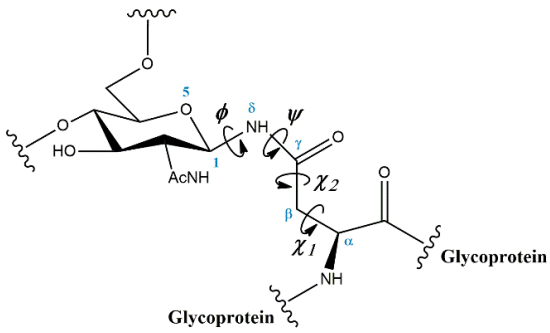


Figure 4.4: Torsion angles adjusted during clash resolution at linkage of PNGs asparagine to GlcNAc. Torsion angles were defined as: φ = O_5-C_1-N_δ-C_γ; ψ = C_1-N_δ-C_γ-C_β; χ_1 = N-C_α-C_β-C_γ; χ_2 = C_α-C_β-C_γ-N_δ.

Table 4.2: PNGs glycosidic linkage torsion angles of the GLYCAM output and MD input structures. Initial and final refer to GLYCAM output and MD input, respectively, with final values given in parentheses where angles were adjusted.

	PNGs	φ initial (final)	ψ initial (final)	χ_1 initial (final)	χ_2 initial (final)
N-domain	**N9**	239.8 (228.5)	166.8	198.6	208.6
	N25	227.8	172.8	177.8	201.5
	N45	267.7	174.1	191.3	178.8
	N82	233.7	165.6 (187.0)	202.3	223.5
	N117	271.6	181.5	194.4	183.9
	N131	234.3	167.8	194.7	221.7
	N289	254.9	176.6	199.8	180.6
	N416	263.9	175.8	189.1	183.6
	N480	255.8	158.9	212.6	201.2
C-domain	**N36**	262.4	172.2	191.3	178.6
	N54	247.7	171.8	192.9	199.0
	N73	267.6 (230.0)	183.4 (192.0)	183.7 (200.0)	174.9 (153.0)
	N119	266.7	169.1	191.5	182.5
	N301	260.9	168.9	195.1	179.5
	N550	229.4 (187.0)	167.5 (190.0)	202.2 (221.0)	210.0 (166.0)

4.4.2 System equilibration during MD simulation

Equilibration of the neutralized and solvated glycoprotein systems was evaluated after each stage of simulation using the *process_minout.perl* and *process_mdout.perl* scripts available in AmberTools 15. Plots showing the variation in energy, temperature, pressure, volume and density for all stages are given in Appendix A17. During the second equilibration stage (unrestrained) the energy, temperature, volume and density of both domains appeared stable (Appendix A17 Figures A14 and A17). These final structures were therefore used as input to the production dynamics simulations. The system parameters were monitored with the *process_mdout.perl* script throughout production to identify any system instabilities. This showed a slight increase in density and further decrease in potential energy and volume over the initial 2500ps of production (Appendix A17 Figures A15 and A18). This likely resulted due to the switch from NVT (constant Number of particles, Volume and Temperature) to NPT (constant Number of particles, Pressure and Temperature) ensemble. The initial 250 frames were therefore regarded as part of the equilibration stage and excluded from analysis. A further 3000 trajectory frames were obtained from 30ns of production dynamics for analyses after equilibration in NPT conditions.

The fluctuation in atomic positions versus the first equilibrated frame was described by the root mean square deviation (RMSD). The protein backbone and all protein heavy atoms showed deviation of about 1Å in both domains, indicating that the protein was stable and no large-scale conformational changes occurred (Figure 4.5). When considering all glycoprotein heavy atoms, however, the RMSD is increased to around 7Å in N-domain and 5Å in C-domain, reflecting the high degree of glycan conformational flexibility.

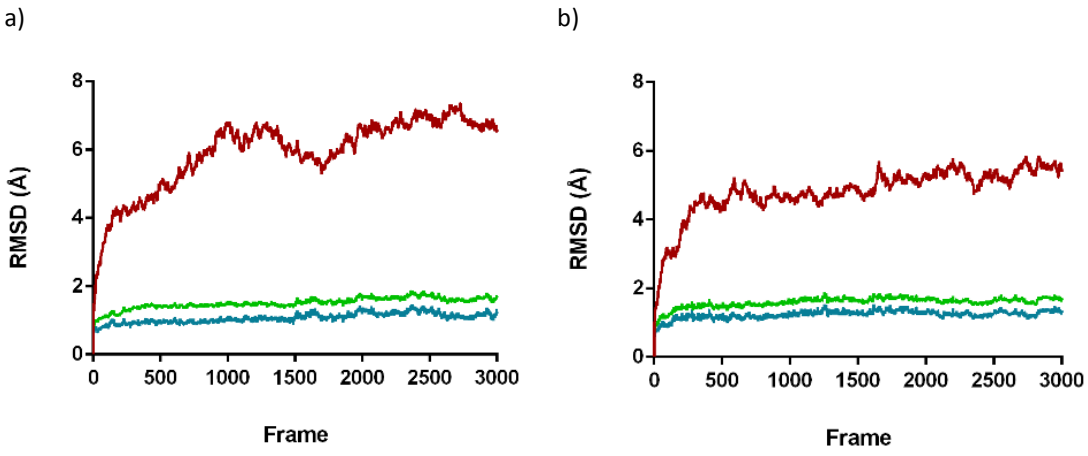


Figure 4.5: Root mean square deviation (RMSD) during MD simulation.
RMSD calculated for a) N-domain and b) C-domain over all glycoprotein heavy atoms, all protein heavy atoms and all protein backbone atoms during 30ns simulation are shown in red, green and blue, respectively.

Since the simulation setup used here differed from that used by Brás *et al* (Brás *et al.*, 2014), the zinc site's behaviour was evaluated by monitoring the distances and angles involved over simulation time. The Zn bonds and angles to HIR, HIT, GLR and WAR were stable in both domains, indicating correct coordination in a tetrahedral geometry (Appendix A17 Figures A19-A22). Additionally, the Zn-bound WAR hydrogens were tightly bound by N-domain E362 and C-domain E348, in agreement with this residue's role in activating the WAR for nucleophilic attack in substrate hydrolysis (Sturrock *et al.*, 2004).

When simulating the dynamic behaviour of ACE in its physiologically active form, one N-domain and two C-domain chloride ions are required since they regulate enzyme activity (Yates *et al.*, 2014). To ensure that these chloride ions did not drift during the simulation, their stability was monitored as well. The N-domain chloride in the second pocket (Cl(II)) and C-domain chloride in the first (Cl(I)) and second pockets (Cl(II)) were stable (Appendix A17 Figures A23 and A24).

4.4.3 Glycan conformational sampling

Glycans have a larger range of motion than the backbone of a protein owing to the flexibility of their *α1-6* glycosidic linkages and the *N*-linked asparagine rotamer. An energy barrier exists which determines the ease of transition between conformational states and

population of each glycan conformation. In glycoproteins, the sampled conformations are further affected by the presence of neighbouring protein and/or glycan atoms (Yang *et al.*, 2017a). If the energy barrier between low energy states is high, transition between states might not be observed using traditional MD approaches. Increased sampling of glycan conformational space was recently described for the HIV gp120/gp41 envelope trimer using enhanced sampling Hamiltonian Replica Exchange with solute tempering and biasing potential (HREST-BP) simulations (Yang *et al.*, 2017a). To assess glycan conformational sampling and biological relevance of the most populated conformations in the current MD simulations, the Asn-GlcNAc, Fuc*α(1-6)*GlcNAc, Man*α(1-6)*Man and NeuAc*α(2-3)*Gal glycosidic torsion angles (Figure 4.6) were monitored over all frames.

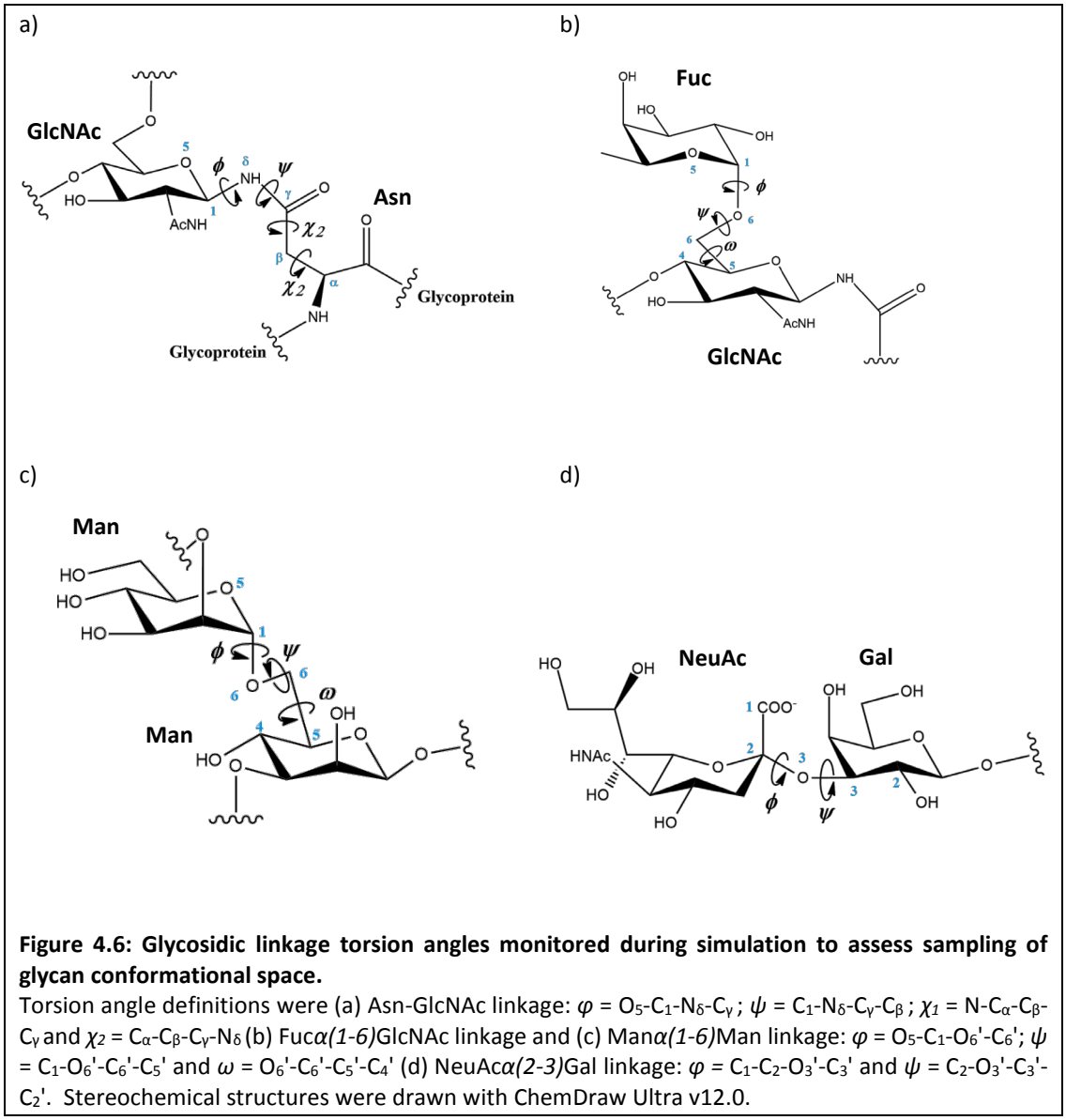


Figure 4.6: Glycosidic linkage torsion angles monitored during simulation to assess sampling of glycan conformational space.
Torsion angle definitions were (a) Asn-GlcNAc linkage: φ = O_5-C_1-N_δ-C_γ ; ψ = C_1-N_δ-C_γ-C_β ; χ_1 = N-C_α-C_β-C_γ and χ_2 = C_α-C_β-C_γ-N_δ (b) Fuc*α(1-6)*GlcNAc linkage and (c) Man*α(1-6)*Man linkage: φ = O_5-C_1-O_6'-C_6'; ψ = C_1-O_6'-C_6'-C_5' and ω = O_6'-C_6'-C_5'-C_4' (d) NeuAc*α(2-3)*Gal linkage: φ = C_1-C_2-O_3'-C_3' and ψ = C_2-O_3'-C_3'-C_2'. Stereochemical structures were drawn with ChemDraw Ultra v12.0.

Overall, the glycosidic torsion angles were comparable between the N- and C-domain (Appendix A18 Tables A3 to A8). The major Asn-GlcNAc linkage conformation had φ ranging between 220° and 270°, a narrow ψ distribution around 170°, a narrow distribution of χ_1 around 180° with some sampling of 280° to 300° and a wide distribution of χ_2 ranging from 60° to 300° (Appendix A18 Tables A3 and A6). These Asn-GlcNAc linkage torsion angles correspond to populations observed upon statistical analysis of protein-glycan linkages in experimental crystal structures (Petrescu *et al.*, 2004).

Similar statistical analyses have also been carried out on the glycosidic torsion angles in glycoproteins (Petrescu *et al.*, 1999). In the current MD simulations the Fuc*α(1-6)*GlcNAc linkage was mostly populated with φ ~280° and ψ ~180° with some minor populations (23% for n1 and 37% for n3) of ψ around ~70° (Appendix A18 Tables A4 and A7). The ω angle displayed conformational transition with a major population (69-99%) of ~50° and a minor population (1-31%) of ~300° to 350° observed for most PNGs. This compares well with the angles of φ = 291.8±9.6°, ψ = 204.1±22.4° and ω = 66.1±14° reported for experimental structures (Petrescu *et al.*, 1999). The Man*α(1-6)*Man linkages showed a narrow distribution of φ ~60° with minor populations (1-5%) of φ ~160° at four N-domain PNGs (Appendix A18 Tables A4 and A7). The ψ angles ranged from ~107° to ~211° and conformational transitions were observed at one C-domain PNGs (ψ = 172° for 15% and ψ = 67° for 85%). For the ω angles, a major population (93-99.97%) of ~50° with conformational transition towards a minor population (0.03-7%) of ~300° was observed at most PNGs. An exception is N-domain N416 which had a major population (85%) of ω ~180° and a minor population (15%) of ω ~60°. This could be the result of favourable protein-glycan interactions and will be described in section 4.4.8. The conformation around this linkage is in line with experimental structures where the major populations have φ = 65.4±9.0°, ψ = 109.1±13.7° or 182.6±5.1° and ω = 66.4±10.2° or 185.0±11.2° (Petrescu *et al.*, 1999).

The reported torsion angles for the terminal NeuAc*α(2-3)*Gal linkage obtained from 14 crystal structures are φ = 68.7±13.6° and ψ = 234.9±15.5° (Petrescu *et al.*, 1999). Again, the current simulations correspond well with a major population (61-99.97%) of φ ~60° and a major population (85-100%) of ψ ~220° (Appendix A18 Tables A5 and A8). Additionally, a second population of φ ~310° was observed to be populated in up to 39% of the trajectory frames. An exception is the *α(2-3)* linkage on the 1-6 arm of C-domain PNGs N54, N73 and N119

where no conformational transition was found, possibly indicative of strong sialic acid interactions as will be described in section 4.4.9. The *α(2-3)* linkage on the 1-3 arm of N-domain PNGs N416 is also an exception since it has a 63% population of φ = 314±11° and only a 37% population of φ = 54±16°.

Further support for the biological relevance of the simulated glycan conformations was derived by comparison of the time-averaged structures to all ACE experimental crystal structures published to date (15 N-domain and 21 C-domain crystal structures). No inferences could be made regarding glycans at N-domain N9, N82 and N131 and C-domain N119, N301 and N550 due to low resolution in these areas. The N25 sidechain and core $GlcNAc_2$ residues aligned well (~3.5Å RMSD for the first GlcNAc) between the average simulated N-domain conformation and two crystal structures (PDB IDs: 2C6F and 2C6N) of fully glycosylated N-domain. Even stronger support was found for the simulated N45 and N416 glycan conformations where the respective $GlcNAc_2Man$ and $GlcNAc_2Man_{1\text{-}3}Fuc$ structures were well-resolved and superimposed with over 10 crystal structures (the first GlcNAc of N45 aligned with a ~3.3Å RMSD to that of PDB IDs: 2C6F, 3NXQ, 4CA6, 4BXK, 5AMC, 4BZS, 5AM8, 5AMB, 4UFA, 6EN5, 2XYD, 5AMA and 5AM9 while the first GlcNAc of N416 aligned with a ~3.3Å RMSD to that of PDB IDs: 5AM9, 5AMB, 4UFB, 4CA6, 5AM8, 5AMC, 4UFA, 3NXQ, 6EN5 and 2XYD). Similarly, the simulated C-domain N36 and N73 conformations were supported by alignment of *N*-linked asparagine sidechains and core GlcNAc, Fuc and Man residues in more than 10 C-domain crystal structures. The time-averaged GlcNAc of N36 aligned with a ~3.3Å RMSD to that of PDB IDs: 2XY9, 2IUL, 4C2O, 2YDM, 4BZR, 4CA5, 4APH, 2IUX, 1O8A, 2IUL, 3BKK, 1UZF and 3BKL while the time-averaged core GlcNAc of N73 aligned with a ~3.4Å RMSD to that of PDB IDs: 2XY9, 4APJ, 2IUL, 3BKK, 3L3N, 3KBL, 4C2Q, 4APH, 4BZR, 4CA5, 4C2P, 4C2R, 4C2N and 4C2O.

Since the simulated glycosidic torsion angles agree with those observed in experimental unrelated crystal structures as well as ACE crystal structures, these dominant glycan conformations are likely of biological significance. Conformations collected over this 30ns timeframe were therefore used for downstream analyses since they showed sampling of physiologically relevant glycan conformations and the RMSD, zinc and chloride ion interactions were stable over this period.

4.4.4 Glycan occupancy of cartesian volume

The cartesian space occupied by each glycan is influenced by its glycosidic linkage flexibility and the resulting frequency of conformational transitions. The VolMap plugin and Volutil package in VMD v.1.9.3 (Humphrey *et al.*, 1996) were used to visualize and quantify the volume occupied by each glycan. For simplicity, from this point onwards glycans at each site will be referred to by a prefix 'n' for N-domain or 'c' for C-domain and their respective position *i.e.* N-domain N9/n1, N25/n2, N45/n3, N82/n4, N117/n5, N131/n6, N289/n7, N416/n8 and N480/n9; C-domain N36/c1, N54/c2, N73/c3, N119/c4, N301/c5 and N550/c6. In line with the initial hypothesis, the N-domain is surrounded by glycans whereas the C-domain subdomain 2 is more exposed (Figure 4.7).

While N-domain n1 has a large range of motion, the volume occupied by n2, n3 and n7 individually is decreased (Figure 4.8). These three glycans form a glycan cluster on subdomain 1, possibly through glycan-glycan interactions. Porcupine plot representation of the direction of dominant glycan motions obtained from PCA illustrated the lateral motion of n2 towards n3 and n7 (Figure 4.9). Principal components 1, 2 and 3 respectively described 28%, 16% and 11% (cumulative 55%) of the overall variation in the N-domain and 32%, 13% and 9% (cumulative 54%) of the overall variation in the C-domain.

A second glycan cluster is formed on subdomain 2 between the stable n5, n9 and the highly dynamic n6 (Figure 4.7). By coinciding at the active site cleft, these clusters could shield the protein from solvent as well as stabilize its closed conformation. N-domain stabilization could also be afforded by the unique n8. This glycan is seen to snugly glide over the protein surface (Figure 4.9) and since it had the lowest volume occupancy, it is likely very stable. Solvent exposure and protein stability can further be affected by the swinging of n4 at the lid's N-terminus on subdomain 1 towards subdomain 2 observed in all three principal components (Figure 4.9).

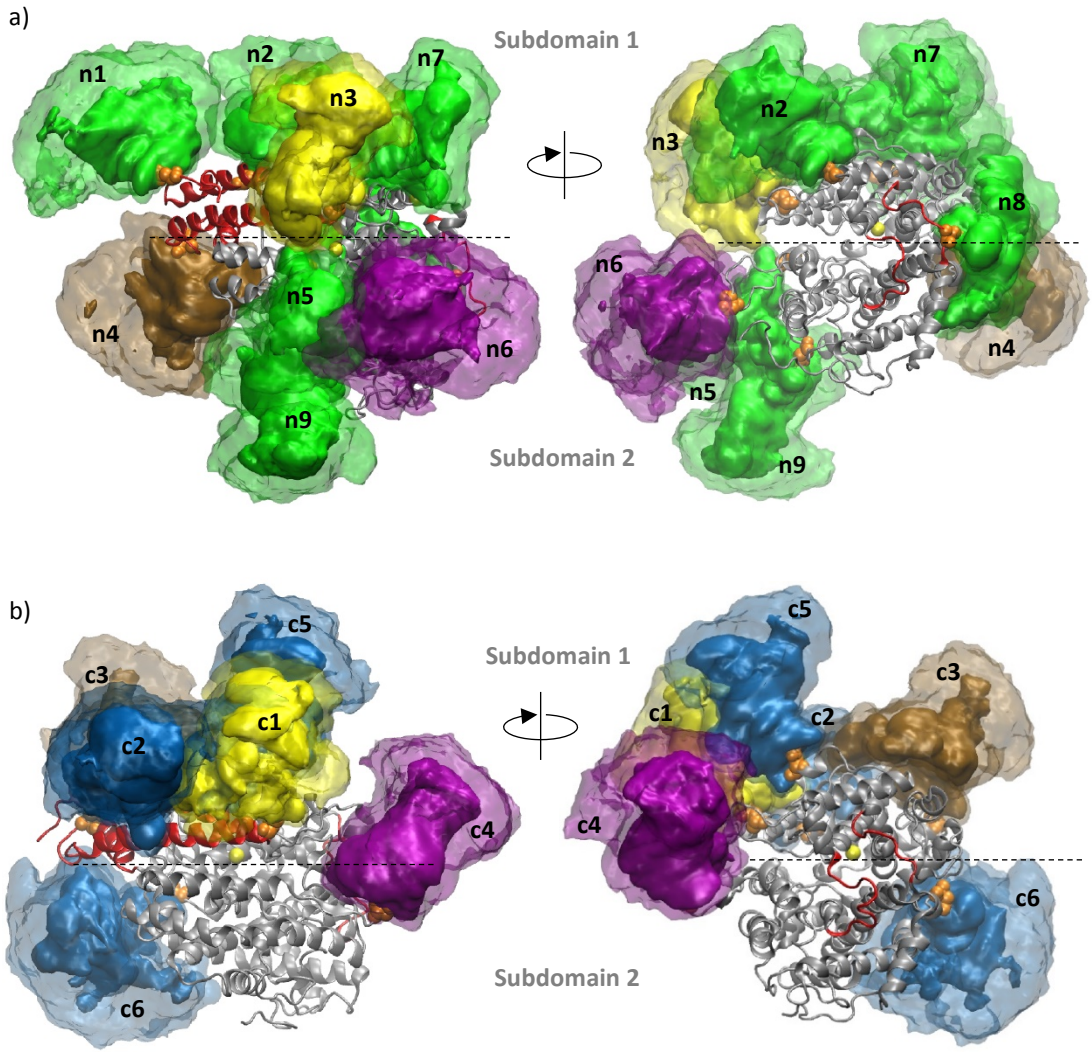


Figure 4.7: Cartesian volume occupied by *N*-glycans.
Isosurfaces represent fractional occupancy volume (normalized to 1 over all frames) contoured at isovalues of 0.05 (opaque) and 0.005 (transparent). Unique a) N-domain and b) C-domain glycans are shown in green and blue, respectively, with the equivalent conserved glycans given in yellow, brown and purple. The catalytic Zn is given as a yellow sphere, *N*-linked asparagines as orange spheres, C-terminal hinges as red loops and the N-terminal lid as red helices. The active site cleft is indicated by a dotted line.

Although regarded as conserved based on sequence alignment, the C-domain c3 counterpart of n4 has a decreased dynamic range and occupies a different region of space (Figures 4.7b and 4.8). Glycan c3 appears to compensate for the lack of a C-domain n7 counterpart by changing direction and approaching the space occupied by this unique N-domain glycan. In turn, the unique C-domain c6 transitions from the protein C-terminus towards the lid's N-terminus to perform the role of N-domain n4 (Figure 4.7). However, while the net motion of n4 described by all three principal components was towards subdomain 2, c6 was more varied and did not appear to be as strongly directed towards subdomain 1 (Figure 4.9).

The conserved glycan c1 has a lower dynamic range than its N-domain counterpart n3 and occupies a slightly different region (Figure 4.7). This is likely due to glycan-glycan interactions upon formation of a glycan cluster on subdomain 1. Although c3 occupies the space of n7, it does not contribute to cluster formation. Instead, the C-domain cluster is formed by c1, c2 and the unique c5 which occupies a space equivalent in location and volume to the unique N-domain n2.

No cluster is formed on the C-domain subdomain 2 due to lack of n5 and n9 counterparts. This likely influenced the location, volume occupancy and direction of motion of the conserved c4 (Figure 4.7-9). Additionally, there is no compensation for lack of a C-domain counterpart to n8. Overall, this increased the C-domain active site cleft and subdomain 2 exposure.

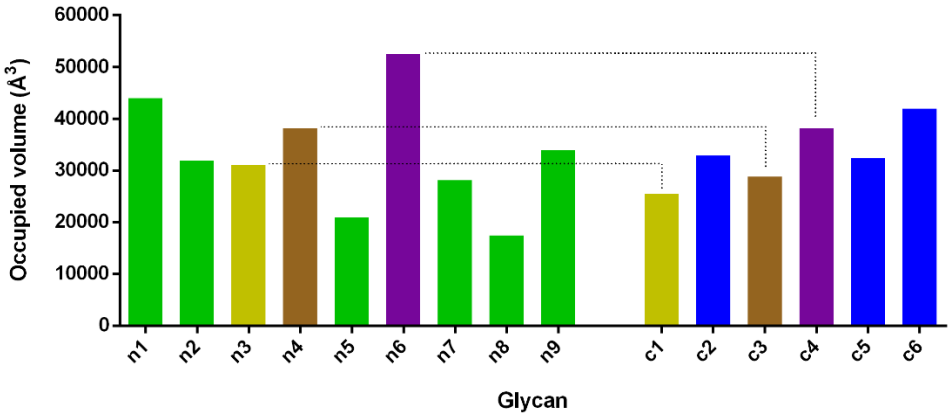


Figure 4.8: Volume occupied by each glycan's heavy atoms calculated over all frames of MD simulation.
Glycans are represented as green bars for N-domain and blue bars for C-domain with the corresponding conserved glycan sites given in yellow, brown and purple.

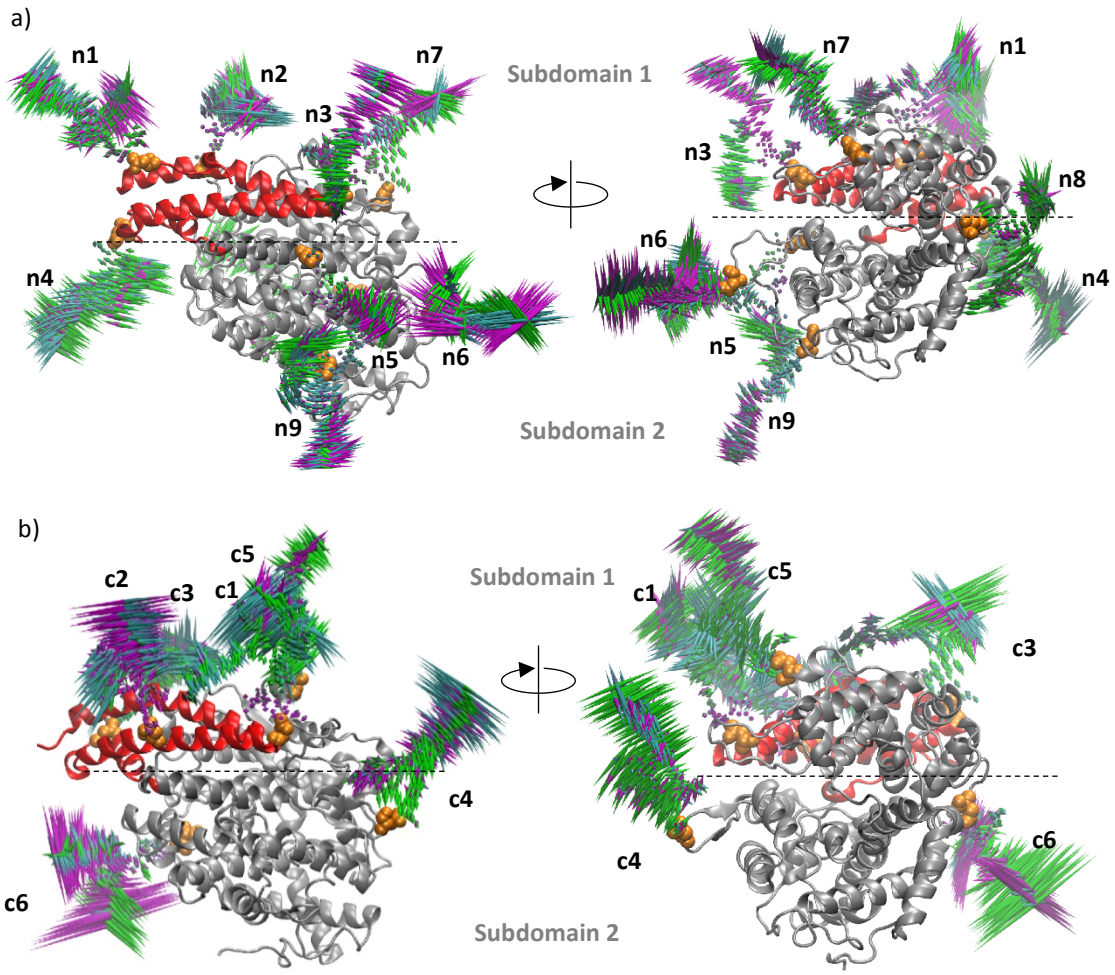


Figure 4.9: Porcupine plot representation of the dominant glycan motions.
Principal components 1 (green), 2 (cyan) and 3 (magenta) respectively are shown for a) N-domain and b) C-domain with porcupine quills illustrating the direction and magnitude of glycan motion. The N-terminal lid is coloured red, *N*-linked asparagine residues given as orange spheres and the active site cleft indicated by a dotted line.

4.4.5 Glycan-glycan interactions

To validate the existence of the proposed glycan clusters and strength of the contacts, radial distribution functions were calculated between glycan pairs. For the most prevalent interactions, hydrogen bond analysis was performed with a 3Å bond length cut-off to identify directly interacting sugars (this did not consider water-mediated bonds or interactions not involving hydrogen bonds).

The minimal n1–n2 and n2–n3 distances between N-domain lid glycans ranged from ~15-50Å and interaction probability was low (Figure 4.10). Although the minimal n2–n7 distance decreased to ~2-50Å, interaction probability remained low. The minimal n3–n7 distance, however, ranged from ~2-40Å and probability of this interaction was increased. The decreased probability of the aforementioned individual interactions might be attributed to competition between n1, n3 and n7 for n2 binding. The subdomain 1 cluster is thus formed by a strong n3–n7 interaction, assisted by interactions of n2 with n7 and, to a lesser extent, n1 and n3. Interaction between n5 and n9 at the N-domain subdomain 2 were highly probable although the minimal distance still extended to ~40Å, reflecting the dynamic nature of glycans. Hydrogen bonds were frequently detected between the n5 sialic acid (0SA) or mannose (2MA) and n9 sialic acid (0SA) or fucose (0fA) residues (Table 4.3). Despite its large range of motion, n6 is seen to interact with n5. The subdomain 2 cluster is thus formed by a strong n5–n9 interaction, assisted by n6. Since the probability of n5 interaction with n3 is rather low, the glycan clusters on each subdomain appear distinct. They could, however, still function synergistically in shielding the N-domain active site cleft.

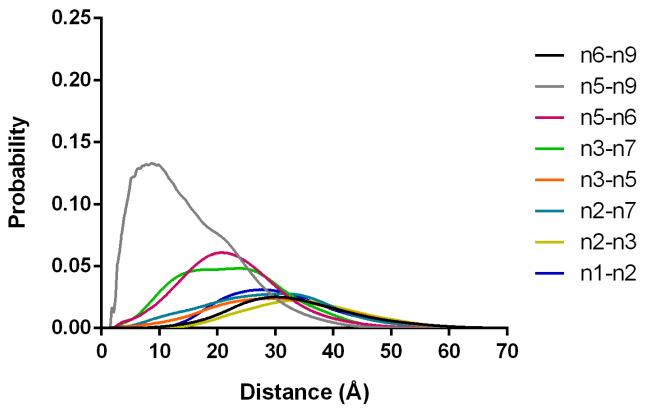


Figure 4.10: Probability of N-domain glycan-glycan interaction (radial distribution function). Interacting glycan pairs are coloured according to the legend.

Table 4.3: N-domain glycan-glycan hydrogen bonds.
Hydrogen bonds present in more than 5% of the simulation are shown with the frequencies indicated in parentheses and acceptor residue names italicized.

Glycan		Donor sugar - acceptor sugar (% frames)
Donor	**Acceptor**	
n5	n9	0SA O4 - UYB O2N (27.7) 0SA O9 - 0SA O4 (24.8) 0SA O7 - 0fA O3 (11.1) 2MA O4 - 0SA O4 (6.1) 2MA O4 - 0SAO8 (5.5)
n9	n5	0fA O2 - 0SA O5N (71.9) 0fA O3 - 0SA O7 (34.6) 0SA O4 - 2MA O3 (9.3) 0fA O3 - 0SA H7O (5.9)

0fA = Fuc, 0SA = NeuAc, 2MA = Man, 3LB = Gal, 4YB = GlcNAc, UYB =1st core GlcNAc, VMB = 1st Man linked to core GlcNAc

Based on the volume occupancy, N-domain n3 and n2 are equivalent to C-domain c1 and c5, respectively (Figure 4.7). Yet, the probability of n2–n3 interaction is markedly lower than that of c1–c5 (Figure 4.11). This is likely because c5 and c1 are more proximal than n2 and n3. Probability of c1–c5 interaction is further increased by decreasing the c1 dynamic range through interaction with c2. Although the volume occupied by C-domain c3 was comparable to N-domain n7 (Figure 4.8), the probability of c1–c3 interaction is markedly lower than n3–n7 (Figure 4.11). Given that c3 is located on the lid's N-terminus, it is likely unable to reach c1 on the lid's C-terminus. The C-domain subdomain 1 cluster is thus formed by strong interactions of c1 to c2 and c5 involving sialic acid (0SA), galactose (3LB), N-acetylglucosamine (4YB), mannose (2MA and VMB) and fucose (0fA) residues (Table 4.4).

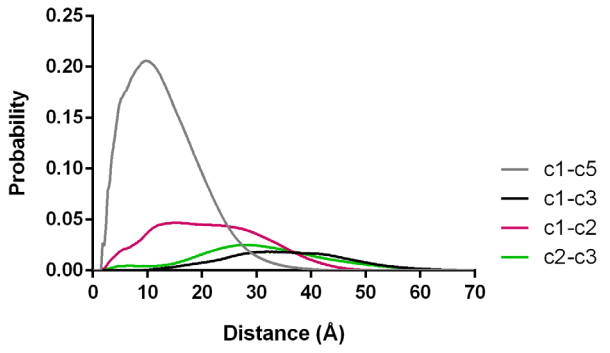


Figure 4.11: Probability of C-domain glycan-glycan interaction (radial distribution function).
Interacting glycan pairs are coloured according to the legend.

Table 4.4: C-domain glycan-glycan hydrogen bonds.
Hydrogen bonds present in more than 5% of the simulation are shown with the frequencies indicated in parentheses and acceptor residue names italicized.

Glycan		Donor sugar - acceptor sugar (% frames)
Donor	**Acceptor**	
c2	c1	0fA H4O - 0SA O1B (5.9) 0fA H3O - 0SA O1B (5.1)
c1	c5	VMB O4 - 3LB O6 (26.6) 2MA O3 - 0SA O1A (26.3) 2MA O4 - 0SA O1B (26.2) 2MA O4 - 0SA O1A (17.2) 4YB O6 - 3LB O2 (15.8) VMB O2 - 4YB O3 (12.9) 2MA O3 - 0SA O1B (11.5) 4YB O6 - 0SA O1B (10.6) 2MA O3 - 3LB O6 (10.1) 0SA O8 - 0SA O1B (9.8) 0SA O9 - 0SA O1A (9.7) VMB O2 - 4YB O2N (7.5) 4YB O6 - 4YB O6 (7.1) 0SA O9 - 0SA O1B (6.9) 0SA O7 - 0SA O1B (5.5) 2MA O4 - 2MA O3 (5.2)
c5	c1	4YB O3 - 2MA O3 (15.3) 3LB O6 - 2MA O4 (14.6) 3LB O6 - VMB O2 (10.4) 0SA O8 - 3LB O2 (9.5) 0SA O8 - 0SA O8 (7.8) 0SA O9 - 4YB O6 (6.9) 3LB 06 - 4YB O6 (5.1)

0fA = Fuc, 0SA = NeuAc, 2MA = Man, 3LB = Gal, 4YB = GlcNAc, UYB =1st core GlcNAc, VMB = 1st Man linked to core GlcNAc

4.4.6 Flexibility of the underlying protein

Dynamic behaviour of the two domains was assessed by comparing the root mean square fluctuation (RMSF) of their backbone protein atoms. Although the overall profile is very similar between N- and C-domain, some differences were noticed which could provide insight into their distinct thermal stabilities (Figure 4.12).

Backbone flexibility of C-domain N-terminal residues 1-100 is markedly increased over N-domain. These residues are located at the lid's N-terminus where differences in glycan occupancies were observed (Figure 4.13). Increased C-domain backbone flexibility was also noticed towards the C-terminal sequence. Interestingly, these residues located not only to subdomain 2 but also hinge region 3 and an active site helix harbouring the zinc-coordinating

glutamate (Figure 4.13). Considering the location of unique glycans in relation to these regions of differential flexibility, glycosylation likely dampened the N-domain's dynamics.

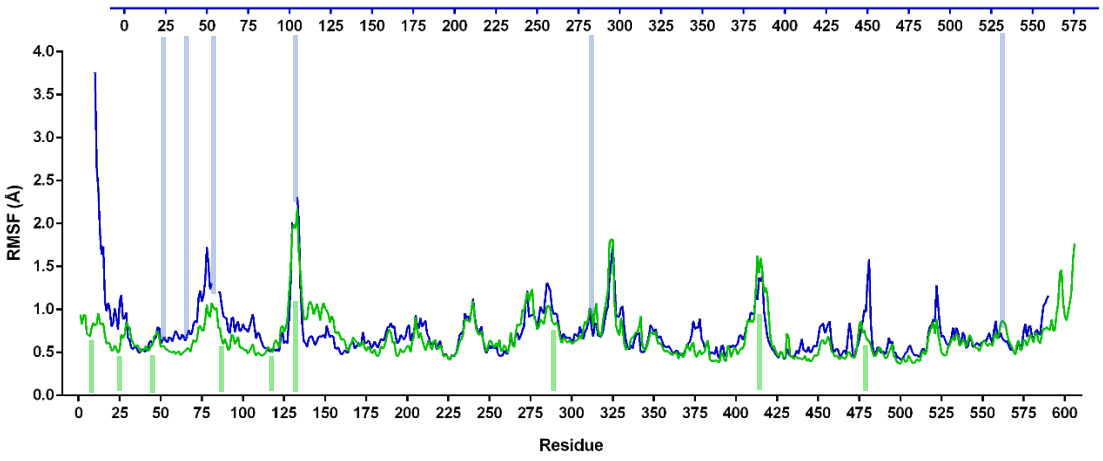


Figure 4.12: Comparison of protein backbone flexibility between the domains.
Root mean square fluctuation (RMSF) of N- and C-domain backbone atoms are given in green and blue, respectively, with lines indicating PNGs locations after structural alignment. N-domain residue numbering is indicated on the x-axis and C-domain residue numbering as a blue ruler above the graph due to the presence of gaps in the sequence alignment.

Two flap-like structures formed via cysteine disulphide bonds were observed on the surface of both domains. The subdomain 1 flap comprises a loop adjacent to the antiparallel beta sheets found parallel to the active site zinc-binding motif (Figure 4.13). Interestingly, the subdomain 2 flap harbours a conserved PNGs (N-domain n6, C-domain c4) and separates two helices lining the active site cleft. Despite conserved glycosylation of this flap, flexibility of the adjacent helix 8 is increased in the N-domain.

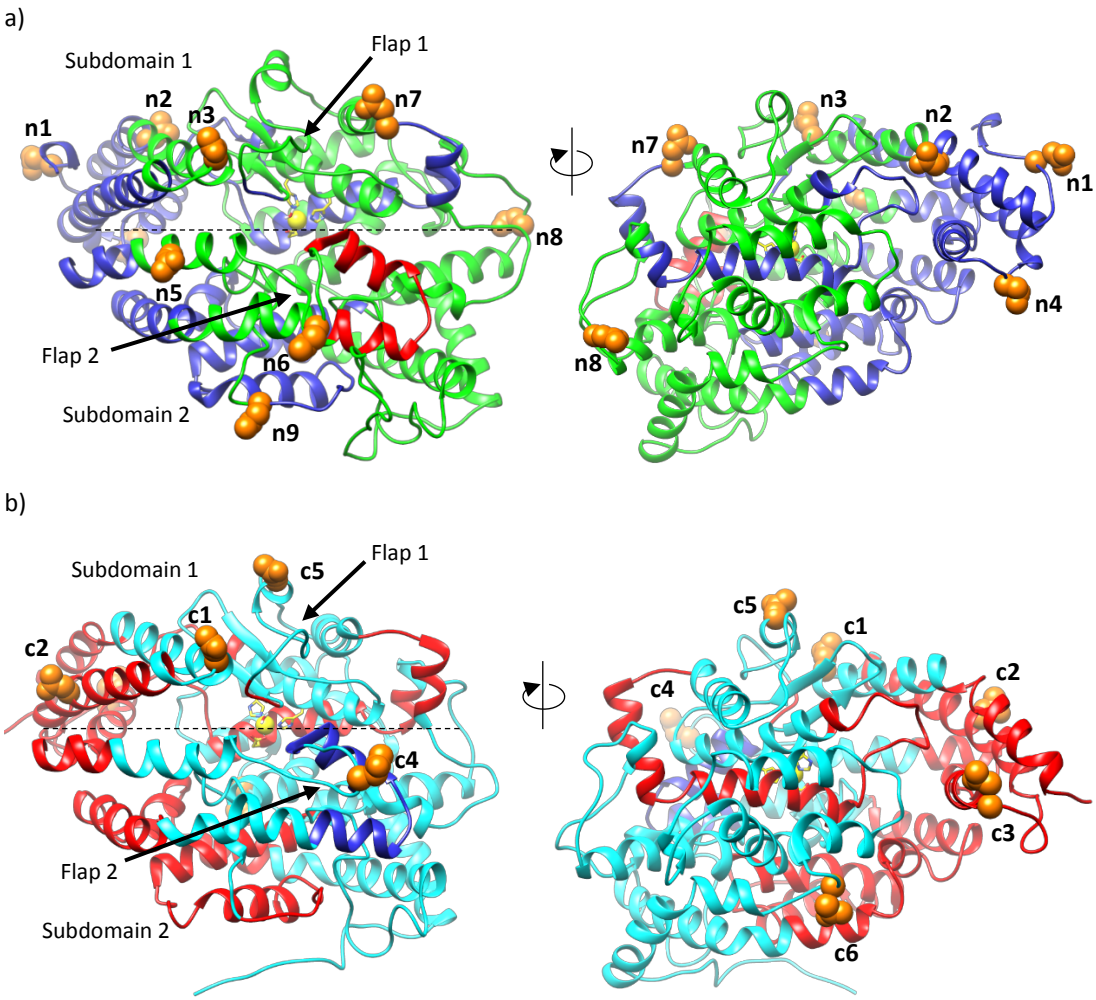


Figure 4.13: Location of residues with differential backbone flexibility.
a) N-domain (in green) and b) C-domain (in cyan) structures are given with regions of increased or decreased RMSF over the opposite domain indicated in red or blue, respectively. *N*-linked asparagines, catalytic Zn and Zn-binding residues are shown as orange spheres, a yellow sphere and yellow sticks, respectively. The active site cleft is indicated by a dotted line and the surface flap structures by arrows.

Differential flexibility of the C-terminal cleft-lining helix 8 is the result of unique protein as well as glycan-glycan interactions. As described in Chapter 2, N-domain hinging to a closed state is mediated by contact between S357 of subdomain 1 and E431 of subdomain 2. This is accompanied by transition from a D354-R350 to D140-R350 salt bridge and a resulting inward motion of helix 8 (Chapter 2). Interaction of n6 with n5/n9 of the subdomain 2 glycan cluster, however, would cause a shift in the subdomain 2 flap towards the N-terminal cleft-lining helix 7 (Figure 4.14). This motion would translate to a shift of helix 8 and formation of an alternate D140 salt bridge to the subdomain 1 flap residue K321. The large number of acidic and basic N-domain flap residues form a chain of alternating salt bridges to encourage

hinging and strengthening of interactions over the active site cleft upon closure. This mechanism is in line with the exceptionally high dynamic range detected for n6 (Figure 4.8).

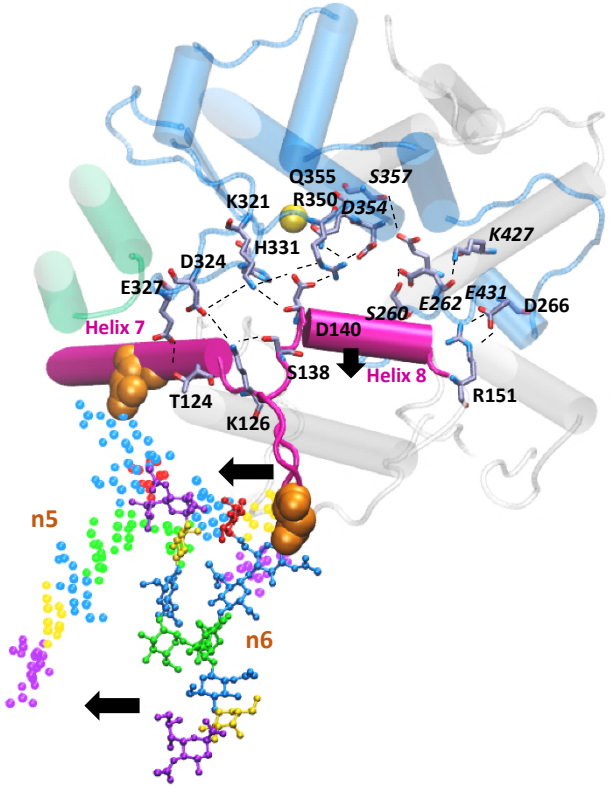


Figure 4.14: Representative N-domain structure illustrating inter-subdomain interactions proximal to n6.
Glycan n5 is shown as dots and n6 as CPK spheres with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015). *N*-linked asparagines are shown as orange spheres, interacting residues as grey sticks and dotted lines, N-terminal lid helices in green and the active site Zn as a yellow sphere. Blue indicates the regions stabilized by N-domain glycans. The cleft-lining helices and subdomain 2 flap is coloured magenta. Residues at the S_2' subdomain interface are italicized. Arrows indicate proposed motion upon interaction of n5 with n6.

In the C-domain, lack of a favourable S_2' subdomain interface increases the stability of helix 8 by decreasing the likelihood of hinging towards a closed S_2' conformation (as discussed in Chapter 2). This is further enhanced by a stable hydrogen bond between H317 of the subdomain 1 flap and E126 of helix 8 (Figure 4.15). Although H317 is conserved, the shorter N-domain counterpart D140 renders this interaction less likely than D140-R350 or D140-K321 salt bridge formation. In the absence of C-domain counterparts to n5 and n9, sideways motion of the subdomain 2 flap would also not occur, in line with the decreased dynamic range of c4 (Figure 4.8).

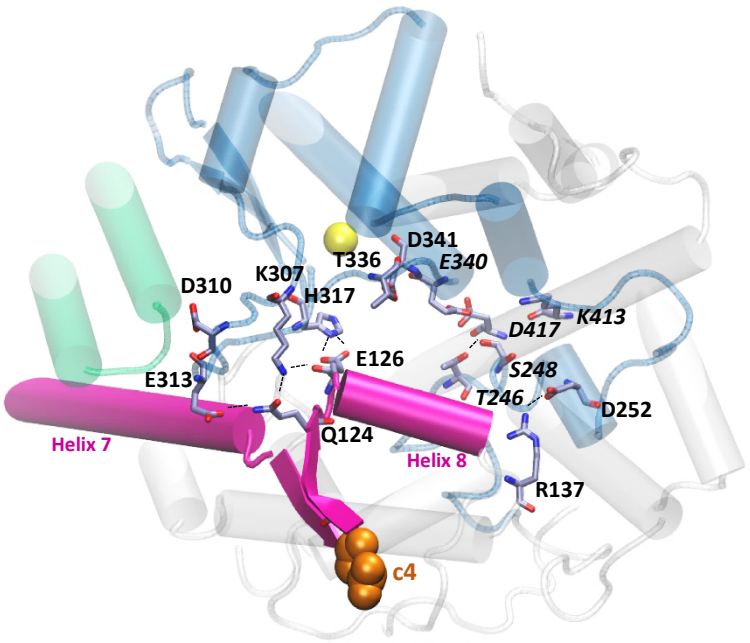


Figure 4.15: Representative C-domain structure illustrating inter-subdomain interactions proximal to c4.
N-linked asparagine is shown as orange spheres, interacting residues as grey sticks, N-terminal lid helices in green and the active site Zn as a yellow sphere. Blue indicates the equivalent regions to those stabilized by glycans in the N-domain. The cleft-lining helices and subdomain 2 flap is coloured magenta. For clarity, c4 atoms are omitted. Residues at the S_2' subdomain interface are italicized.

4.4.7 Glycan interactions with the N- and C-domain

The mechanisms whereby glycans increased protein stability and the contribution of these towards thermal stability was investigated by analysing glycan-protein hydrogen bond frequencies. Only hydrogen bonds which were present in more than 5% of the simulation were considered. To visualize these hydrogen bonds, the most prevalent glycan conformations were extracted using clustering analysis.

Apart from n1, n6 and n9, all N-domain glycans formed numerous hydrogen bonds to the protein (Appendix A19 Figures A25-A26 and Tables A9-A10). Frequent hydrogen bonds were also observed between the C-domain and glycans c1, c3 and c6 (Appendix A19 Figures A27-A28 and Tables A11-A12). Although the occupancy volumes suggested that C-domain c5 plays a similar role to N-domain n2, the protein interaction to c5 was markedly lower than n2.

4.4.8 The role of glycosylation in N-domain thermal stability

As discussed in Chapter 1 sections 1.3.2 and 1.3.3, glycans n2, n3, n7, n8 and n9 are involved in N-domain folding and thermal stability. The n3 and n7 fucose, galactose and sialic acid residues were observed to hydrogen bond residues of N-domain subdomain 1, all of which are unique except E49 and R344 (Figure 4.16). The strength of these interactions is exemplified by their presence in a large proportion of the simulation (Appendix A19 Figures A25-A26 and Tables A9-A10). Apart from the glycan interactions formed to n3 in the subdomain 1 cluster, this glycan bridges over the tips of parallel helices 2 and 3, thus ensuring their concerted motion and stabilizing the lid's C-terminus (Figure 4.16). This, in turn, might indirectly stabilize the adjacent 317-352 cysteine disulphide-containing active site loop. Direct stabilization of this loop is achieved via n7 sialic acid and galactose hydrogen bonding to the conserved R344 and unique E315 and G316. Furthermore, stretching of n7 over helix 17 to bind R344 could stabilize the loop on which it resides (hinge region 2) as well as the adjacent active site loop harbouring Q259. Stability in this region is important since Q259 forms part of the oxyanion hole responsible for binding the substrate's C-terminal carboxylate.

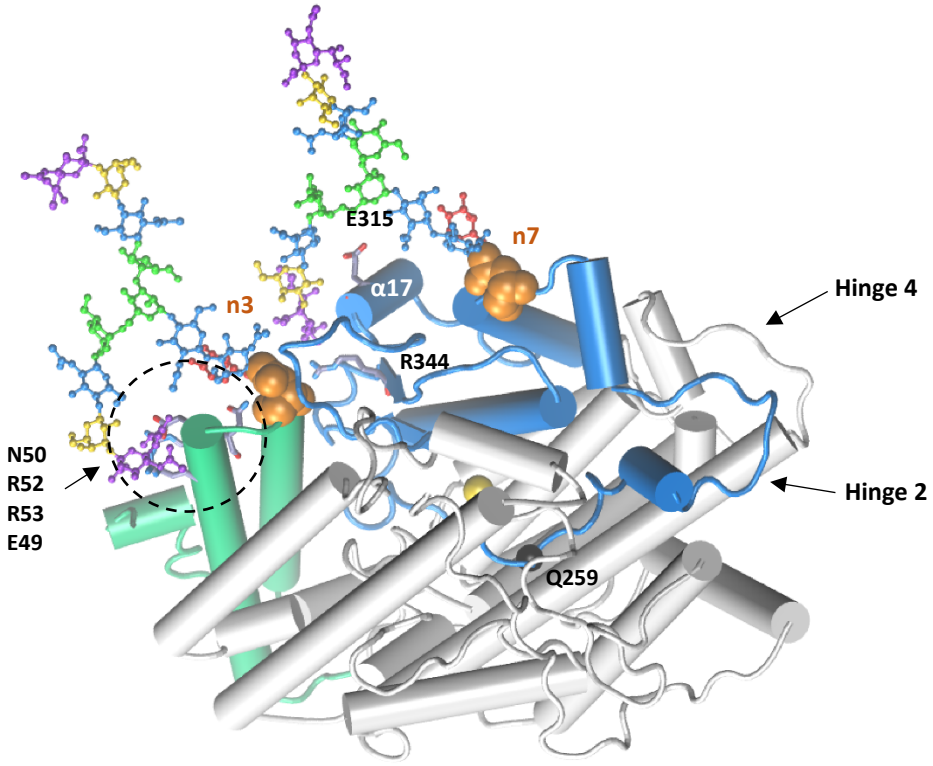


Figure 4.16: Representative N-domain structure showing glycan-protein hydrogen bonding at n3 and n7.
Glycans are shown as CPK spheres with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015), *N*-linked asparagines as orange spheres, interacting N-domain residues as grey sticks, N-terminal lid helices (helices 1-3) in green, regions proposed to be stabilized by glycan interactions in blue, active site Zn as a yellow sphere and the conserved oxyanion residue Q259 C_α as a black sphere. All interacting residues except E49 and R344 are unique. Helix 17 is labelled as α17.

Active site loop stabilization is further augmented by the presence of n2. Apart from n2's role in formation of the subdomain 1 glycan cluster, it forms various stable hydrogen bonds to the protein (Appendix A19 Figures A25-A26). The unique lid residues Q22 and Q30 adjacent to n2 were bound in more than 10% of the simulation (Appendix A19 Tables A9-A10). Interestingly, very strong interactions were observed between the core N-acetylglucosamine and the unique R340 located on the cysteine disulphide-containing active site loop (Figure 4.17). Through a sweeping motion of the n2 1-3 arm, transient sialic acid interactions also occurred to the conserved Q371, R380 and unique C-terminal residues R541 and Q545. This motion might not only stabilize the lid but also assist in its closure at the N-terminal cavity. Since R380 resides on a loop described as hinge region 3 (Watermeyer *et al.*, 2006) parallel to the cysteine disulphide-containing active site loop, binding to n2 might affect hinging and confer stability to the adjacent Zn-coordinating helix 18. Further N-domain C-terminal stabilization was achieved by very strong hydrogen bonds to the unique E518, E522, V560, Q568, Q575 and Q579 residues as n8 wrapped around the protein surface (Figure

4.17). Since n8 is located on hinge region 4 these interactions might not only affect thermal stability but also protein hinging to an open state.

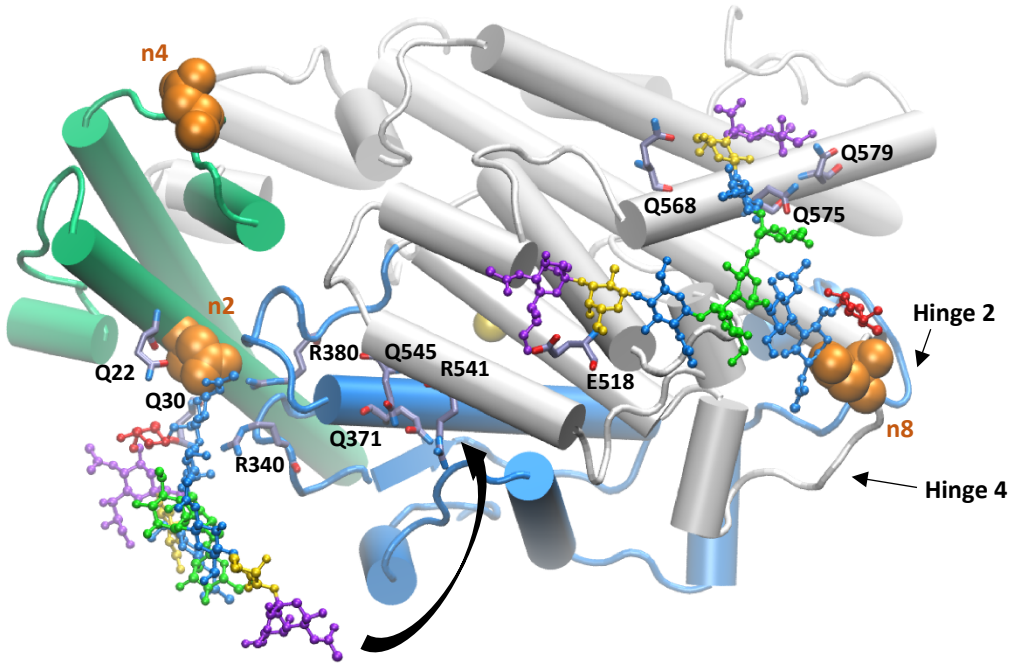


Figure 4.17: Representative N-domain structure illustrating glycan-protein hydrogen bonding at n2 and n8.
Glycans are shown as CPK spheres with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015), *N*-linked asparagines as orange spheres, interacting N-domain residues as grey sticks, N-terminal lid helices in green, regions proposed to be stabilized by glycan interactions in blue, active site Zn as a yellow sphere and the conserved oxyanion residue Q259 C_α as a black sphere. All interacting residues except Q371 and R380 are unique.

Analogous to n3, n4 stabilizes the lid's N-terminus by forming a bridge between helices 2 and 3 upon binding to the unique R89 (Figure 4.18). It further contributes to C-terminal protein stabilization and closure of the N-terminal cavity by interacting with the unique N203. Since n4 is located on hinge region 1, transient interaction to subdomain 2 might drive the lid's tipping motion and explain this glycan's large dynamic range.

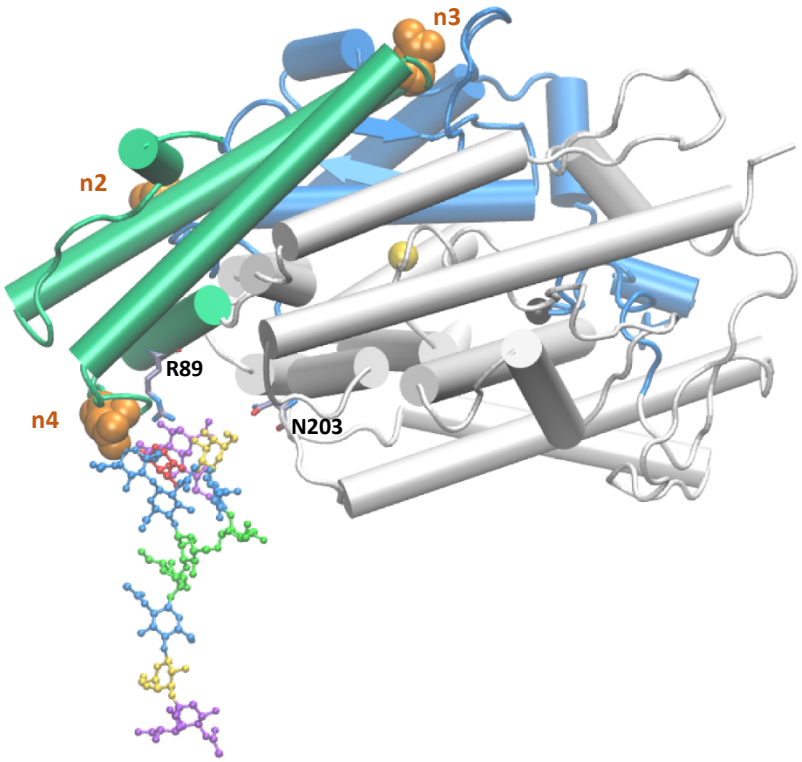


Figure 4.18: Representative N-domain structure illustrating glycan-protein hydrogen bonding at n4. Glycans are shown as CPK spheres with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015), *N*-linked asparagines as orange spheres, interacting N-domain residues as grey sticks, N-terminal lid helices in green, regions proposed to be stabilized by glycan interactions in blue, active site Zn as a yellow sphere and the conserved oxyanion residue Q259 C_α as a black sphere.

Glycans n2, n3 and n7 thus synergistically stabilize the cysteine disulphide-containing active site loop while synergism between n2, n4 and n8 contributes to C-terminal stabilization. Furthermore, synergism between n2, n4, n7 and n8 has the potential to stabilize the entire protein by affecting hinging at regions 3, 1, 2 and 4, respectively.

Although n9 rarely interacts with protein residues, it might contribute to protein stabilization through formation of the subdomain 2 glycan cluster. The strong n5 – n9 glycan interactions restrict n5's dynamic range and strengthen its binding to H483 and the N480 delta nitrogen (Figure 4.19). Cluster formation likely also forms a barrier to prevent the N480-containing loop from exiting the protein. This long loop extends from the surface into the active site where it terminates with Y498, an oxyanion hole-forming residue critical for binding the substrate's C-terminal carboxylate.

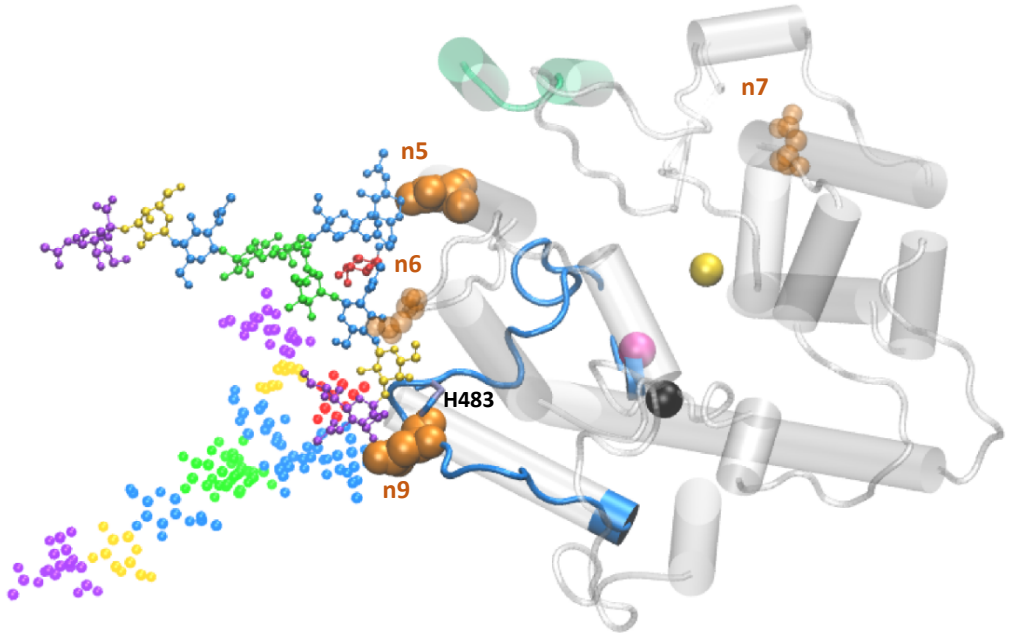


Figure 4.19: Representative N-domain structure illustrating glycan-glycan and glycan-protein hydrogen bonding at n5 and n9.
Glycans are shown as CPK spheres with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015), *N*-linked asparagines as orange spheres, interacting N-domain residues as grey sticks, N-terminal lid helices in green, regions proposed to be stabilized by glycan interactions in blue, active site Zn as a yellow sphere and the conserved oxyanion residues Q259 C_α and Y498 C_α as black and pink spheres, respectively. Only n5 and n9 glycan atoms are shown for clarity.

Experimentally, it has been observed that the combined loss of n7, n8 and n9 leads to protein misfolding and that folding can be restored by reintroduction of either of these in the presence of one or more N-terminal glycans (Anthony *et al.*, 2010). Glycans n7, n8 and n9 together, however, allow protein folding in the absence of N-terminal glycans. The molecular dynamics simulations provided insight into the basis for this specific glycan requirement. During protein folding and processing, combined stabilization of the lid, active site loops, hinges as well as the protein C-terminus is crucial. This can be achieved by hydrogen bonding to a single glycan at each of these regions. Synergistic glycan interactions, however, are required to maintain the protein fold and prevent unravelling of critical substrate-binding loops at elevated temperatures. It has been shown experimentally that n2, n3, n7, n8 and n9 synergistically govern thermal stability (Anthony *et al.*, 2010). This agrees with the *in silico* data presented here which revealed their role in stabilizing hinge regions 2-4, two crucial substrate-binding loops and a cysteine disulphide-containing active site loop proximal to the Zn binding site. The latter appears to be the most labile since its stabilization necessitates direct n2 and n7 binding as well as reinforcement by strong n3-n7 glycan interactions.

4.4.9 The role of glycosylation in C-domain thermal stability

Analogous to what was observed at the N-domain n3, c1 stabilized the lid's C-terminus by forming a bridge between helices 1 and 2 upon sialic acid interaction to E25, Q47, K48 and Q51 (Figure 4.20). Although c5 and N-domain n2 appeared to occupy similar regions of space, c5 is strongly bound to c1 and only rarely contacts the protein at lid residues E28 and W31. Glycan c5 thus did not stabilize hinge region 3 and its more distal location from hinge region 2 likely prevents stabilization of the loop harbouring oxyanion hole residue Q245. Through formation of the subdomain 1 glycan cluster it could, however, stabilize the cysteine disulphide-containing active site loop on which it resides.

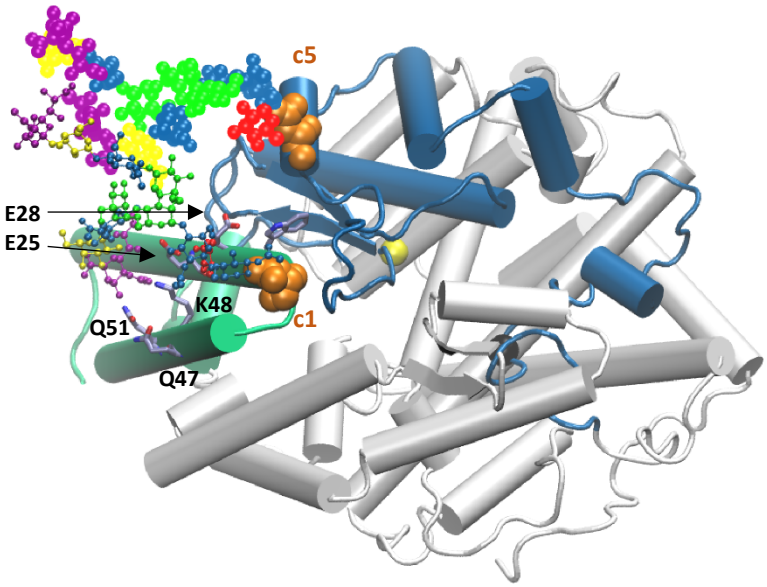


Figure 4.20: Representative C-domain structure illustrating glycan-protein hydrogen bonding at c1 and c5.
Glycan c1 is shown as CPK spheres and c5 as dots with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015). *N*-linked asparagines are shown as orange spheres, interacting C-domain residues as grey sticks, N-terminal lid helices in green, active site Zn as a yellow sphere and the conserved oxyanion residue Q245 C_α as a black sphere. Blue indicates the equivalent regions to those stabilized by glycans in the N-domain.

Along with c1, c2 of the subdomain 1 cluster was also observed to hydrogen bond to helix 1 and 2 lid residues (Figure 4.21). These sialic acid interactions to E13, K58 and Q62 are, however, transient in nature and unlikely to offer much stability to the lid's N-terminus (Appendix A19 Figures A27-A28 and Tables A11-A12). Although c3 and N-domain n4 are both located at the lid's N-terminus (hinge region 1), c3 occupies a region of space resembling N-

domain n7 more than n4. By forming a strong sialic acid interaction to N325, c3 stabilizes the cysteine disulphide-containing active site loop and appears to compensate for lack of a C-domain n7 equivalent (Figure 4.21). However, interaction to D360 was only transient and hinge 3 stabilization thus inferior to that of n7. In forming these interactions, c3 is also unable to sweep across the N-terminal cavity and bind subdomain 2.

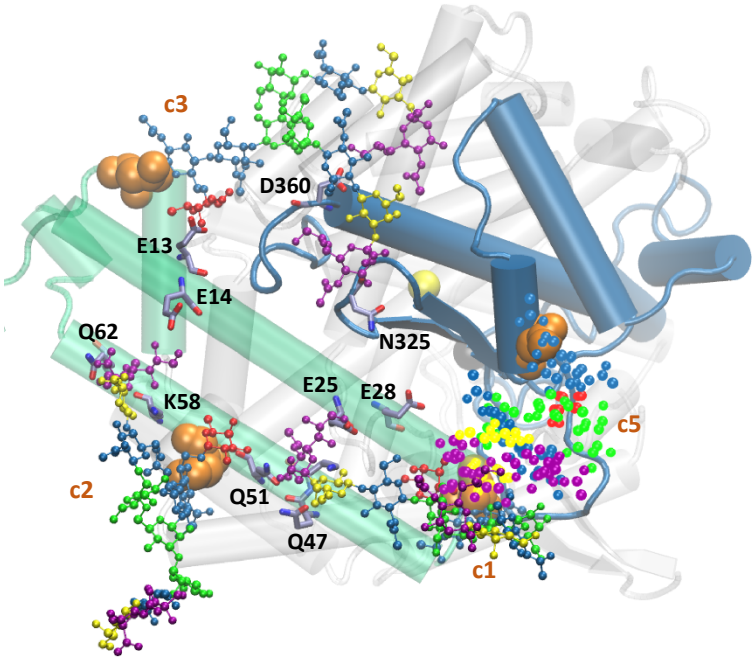


Figure 4.21: Representative C-domain structure illustrating glycan-protein hydrogen bonding at c2 and c3.
Glycans c1, c2 and c3 are shown as CPK spheres and c5 as dots with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015). *N*-linked asparagines are shown as orange spheres, interacting C-domain residues as grey sticks, N-terminal lid helices in green, active site Zn as a yellow sphere and the conserved oxyanion residue Q245 C_α as a black sphere. Blue indicates the equivalent regions to those stabilized by glycans in the N-domain.

Occupancy volume analysis suggested that c6 could perform the same role as N-domain n4. Although this glycan formed very strong hydrogen bonds to the protein (Appendix A19 Figures A27-A28 and Tables A11-A12), these were between the core N-acetylglucosamine and the proximal S192, D196 and M551 residues of subdomain 2 (Figure 4.22). The lack of interactions between c6 and the lid's N-terminus could be attributed to the C-terminal location of N550. This likely contributes to the C-domain's increased backbone flexibility of subdomain 2 and the lid's N-terminus. Furthermore, this likely renders the C-domain N-terminal cavity more exposed to solvent.

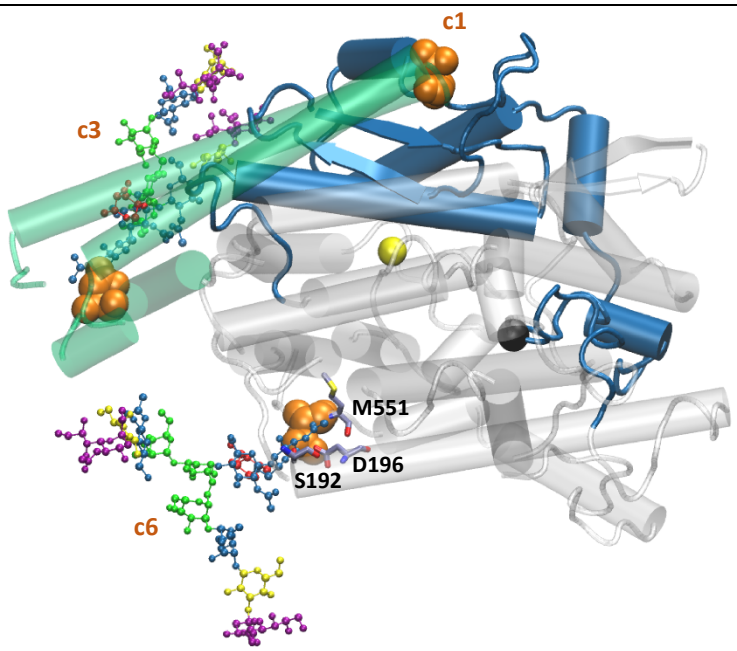


Figure 4.22: Representative C-domain structure illustrating glycan-protein hydrogen bonding at c6. Glycans are shown as CPK spheres with standard Symbol Nomenclature for Glycans (SNFG) colouring (Varki *et al.*, 2015). *N*-linked asparagines are shown as orange spheres, interacting C-domain residues as grey sticks, N-terminal lid helices in green, active site Zn as a yellow sphere and the conserved oxyanion residue Q245 C_α as a black sphere. Blue indicates the equivalent regions to those stabilized by glycans in the N-domain.

Experimentally, it has been observed that C-domain thermal stability is drastically decreased upon removal of c5 and c6 (O'Neill *et al.*, 2008). This decrease is slightly augmented upon additional removal of c2 and c4. As seen for the N-domain, stabilization of the cysteine disulphide-containing active site loop thus appears most crucial for protein stability. This *in silico* study is the first investigation into the basis for differential thermostability of the two ACE domains and, importantly, suggests that the N-domain is more resistant to thermal degradation than C-domain due to an increase in oxyanion hole, hinge, lid and subdomain 2 stabilizing interactions. Unique n2 and n4 N-domain interactions and cluster formation between n5, n6 and n9 likely increased the lid and subdomain 2 stability over C-domain. Unique N-domain interactions with n4, n7 and n2 further increased the stability of hinges 1-3 while hinge 4 was equally flexible in both domains.

N- and C-terminal protein stability, however, also depends upon the nature of their comprising residues. In the N-domain, hydrogen bonds are observed between unique polar residues lining the active site cleft (Figure 4.23a). By stretching across the cleft, these interactions form an 'ionic zipper' between the two subdomains. The C-domain

counterparts, in contrast, are mostly hydrophobic with only one potential hydrogen bond detected (Figure 4.23b). This likely contributes to the C-domain's increased flexibility and ultimately could increase its active site exposure.

Glycan forcefields have been validated by comparing the results of solvated glycans obtained from simulation to that obtained by experimental nuclear magnetic resonance. At present, there is no means to adequately validate the extent to which glycan-glycan or glycan-protein interactions predicted by the forcefield are populated in reality. Since the simulated interactions of glycosylated N- and C-domain ACE presented here agree with melting temperature and protein expression experiments, it nevertheless provides a qualitative molecular basis for the experimentally observed involvement of *N*-linked glycosylation in ACE thermal stability and folding, although it might not reflect the population of these interaction precisely.

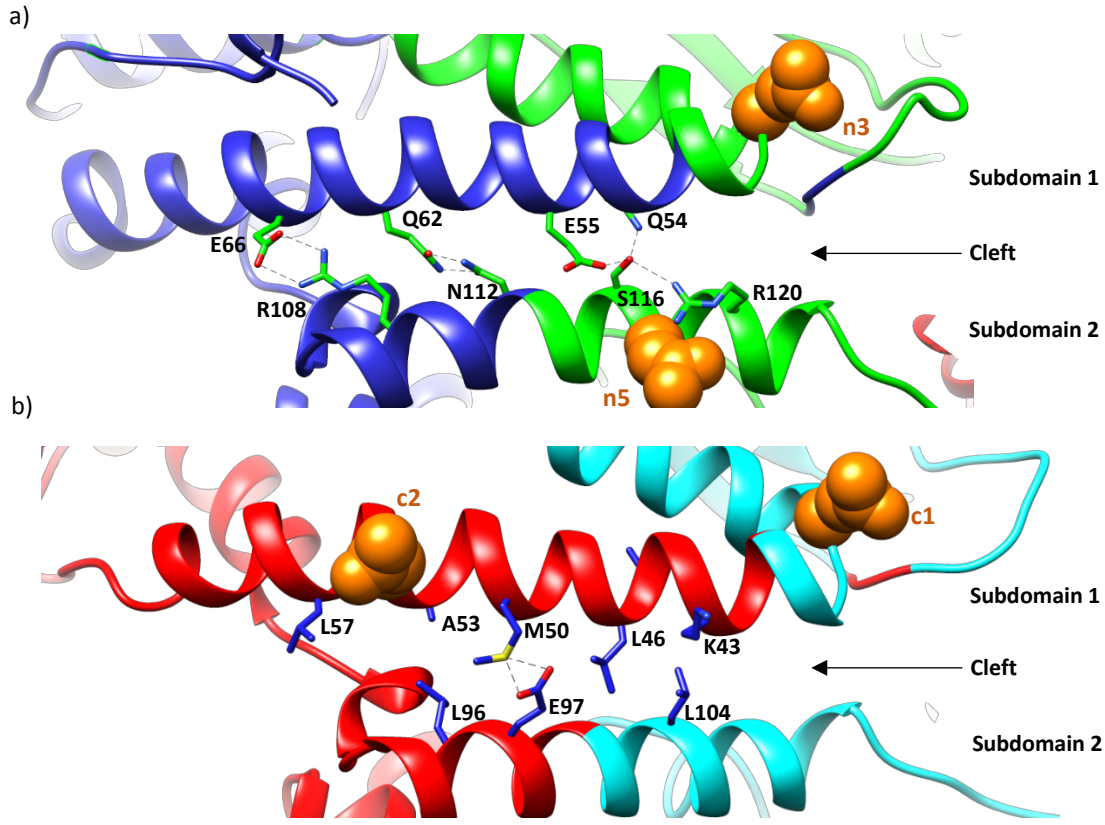


Figure 4.23: Contribution of unique N- and C-domain cleft-lining residues towards protein stability. a) N-domain and b) C-domain are given in green and cyan, respectively. The region with lower N-domain backbone flexibility is coloured blue in a) and red in b). The active site cleft is indicated by an arrow and hydrogen bonds by dotted lines.

4.4.10 Implications for a minimally glycosylated sACE glycoform

As discussed in Chapter 1 section 1.4.2, attempts at obtaining a minimally glycosylated sACE form have thus far failed due to misfolding or inter-domain linker cleavage of N>Q mutant sACE proteins. The *in silico* data presented here offers insight into this phenomenon. A sACE model was constructed from representative fully glycosylated N- and C-domain conformations. The N-domain was orientated normal to the molecule's long axis while the C-domain was inclined by ~60°, as suggested by low resolution electron microscopic 3D reconstruction of porcine sACE (Chen *et al.*, 2010). In accordance with their reconstruction, the domains were separated by 2.5nm.

Glycans which could potentially shield and/or stabilize the linker were then identified to propose a minimally glycosylated sACE variant resistant to linker proteolysis. In this sACE model, the N-domain hinges 2 and 4 and its C-terminal helices are proximal to the C-domain N-terminal lid (Figure 4.24), reminiscent of the arrangement observed via X-ray crystallography between molecules of neighbouring N-domain asymmetric units (discussed in Chapter 2 section 2.6.12). Instability at any of these regions could destabilize the inter-domain linker and possibly increase its exposure by altering the relative domain orientation. The protein might then be susceptible to cleavage into single domains. Glycan n4 and n8 can function to stabilize the N-domain C-terminal helices and hinge 4. The adjacent hinge 2 can be stabilized by n7. This would also ensure folding and thermal stability by binding to flap 1. Although fifteen minimally glycosylated N-domain (n3-n8-n9) crystal structures illustrate that equivalent stabilization of the lid, flap 1 and hinge 2-flanking helices can be obtained by binding of polyethylene glycol molecules, glycan n7 would nevertheless be required during protein expression. The glycan cluster formed by n5 and n9 is proposed to be essential since n9 resides on a very long loop. Lack of stabilization by this cluster might be transferred to the neighbouring helices (coloured light blue on Figure 4.24) and destabilize hinge 4 or the C-terminal helices (coloured black on Figure 4.24). Experimentally, a 70% decrease in linker cleavage was observed upon introduction of n6, and thus glycan clustering, to an n2-n3-n8-n9-c1-c3 sACE (Nkoe, 2014). Based on the MD simulations, this decrease could be augmented by rather including n5 since its n9 interaction is stronger.

On the C-domain, the glycan cluster formed by c1 and c5 is necessary to stabilize the lid's C-terminus. Lid stability is further ensured by stabilizing the adjacent flap 1 via c3. Analogous to n7, this is likely also required for correct folding in cell culture. Furthermore, there is the

possibility of inter-domain glycan clustering via c1/c5–n7 and c3–n8 or inter-domain glycan to protein contacts. It is therefore proposed that the minimal glycan requirement for a stable sACE protein for structure determination via X-ray crystallography is n4,5,7,8,9 c1,3,5 (53% of the wild-type carbohydrate content).

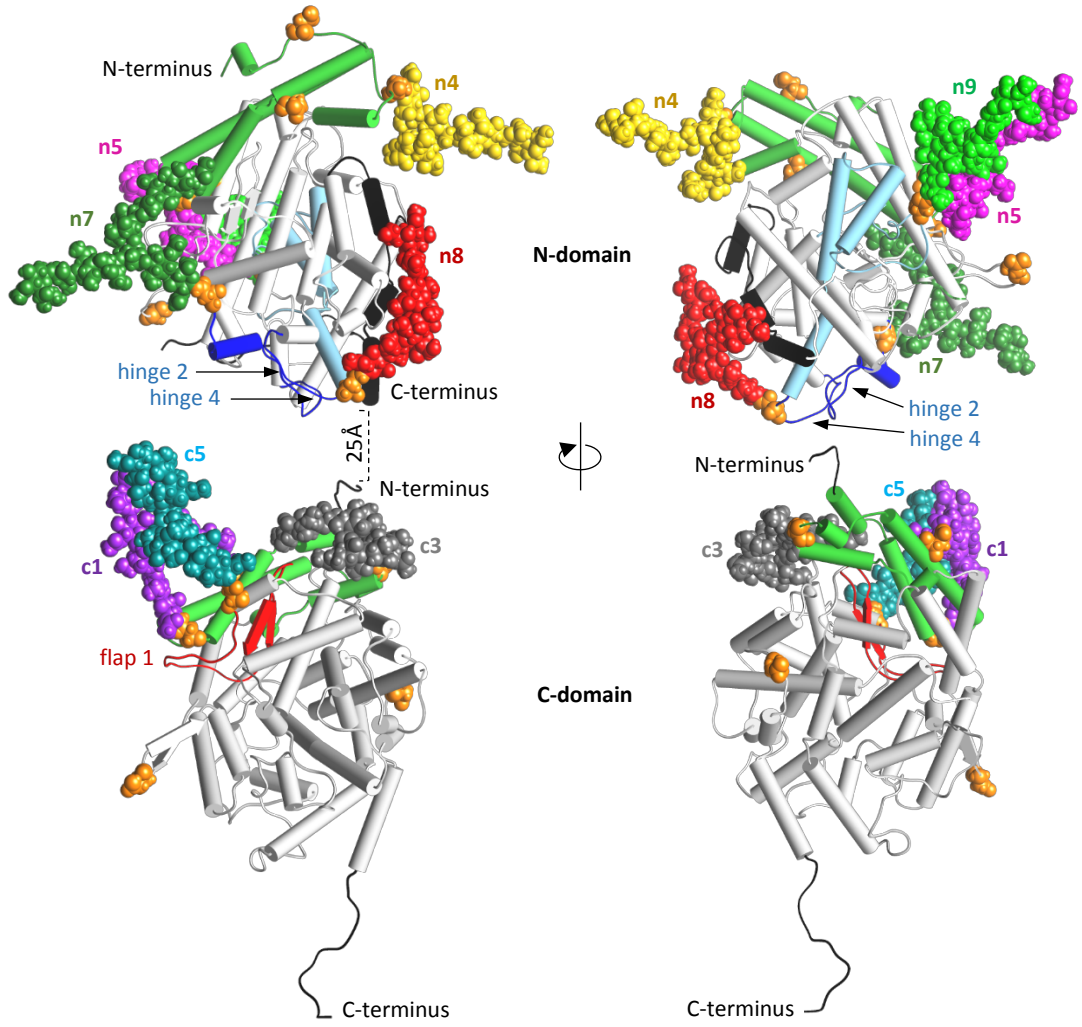


Figure 4.24: Model of a proposed minimally glycosylated sACE variant.
The N-terminal lid regions are shown in green, N-linked asparagines as orange spheres and the stalk as a black coil (after modelling it into an extended conformation). The inter-domain linker was omitted since it packed onto the protein surface in the truncated N-domain structure. The N-domain C-terminal helices are coloured black, hinges 2 and 4 dark blue and the helices connecting n8 to the n5/n9 cluster light blue. Glycan atoms of the proposed minimally glycosylated sACE are shown as spheres.

4.4.11 The role of glycosylation in regulating oxidative inactivation

To investigate the molecular mechanism underlying the C-domain's increased susceptibility to oxidative inactivation, Caver Analyst v1.0 (Kozlikova *et al.*, 2014) was used. All tunnels providing access to the active site Zn from bulk solvent were computed dynamically for the two domains over the last 100 trajectory frames (1ns). Both domains displayed tunnels in the prime and non-prime subsites (Figure 4.25). The C-domain, however, had a greater number of distinct tunnel clusters, particularly at the non-prime lid region.

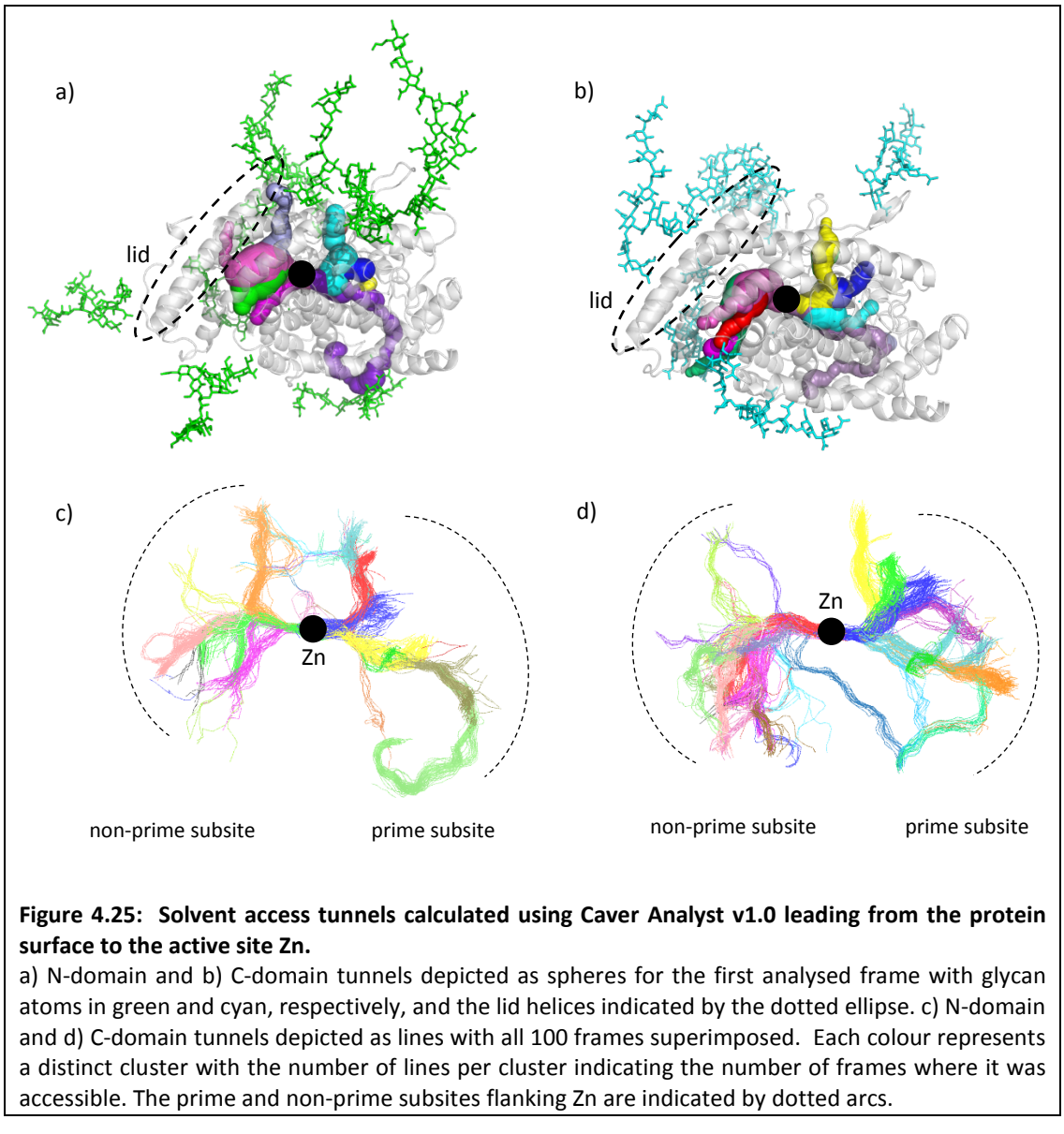


Figure 4.25: Solvent access tunnels calculated using Caver Analyst v1.0 leading from the protein surface to the active site Zn.
a) N-domain and b) C-domain tunnels depicted as spheres for the first analysed frame with glycan atoms in green and cyan, respectively, and the lid helices indicated by the dotted ellipse. c) N-domain and d) C-domain tunnels depicted as lines with all 100 frames superimposed. Each colour represents a distinct cluster with the number of lines per cluster indicating the number of frames where it was accessible. The prime and non-prime subsites flanking Zn are indicated by dotted arcs.

While the N-domain had one major access tunnel between the subdomain 1 and 2 flaps (cyan spheres in Figure 4.25a), additional tunnels formed in the C-domain at the S_2' subdomain interface (Figure 4.26). When aligning the two domains it was observed that unique S_2'

residues of the N-domain overlapped with the C-domain tunnels. These residues thus limit solvent access between the subdomains in the N-domain by interacting via hydrogen and ionic bonds (as detailed in section 4.4.6, Figure 4.14). In the C-domain, replacement of these gating residues by smaller and/or hydrophobic counterparts allowed tunnel formation and thus increased solvent access. Increased ROS influx via this subsite can lead to oxidation of the Zn-coordinating histidine residues as well as the substrate-binding oxyanion hole tyrosine.

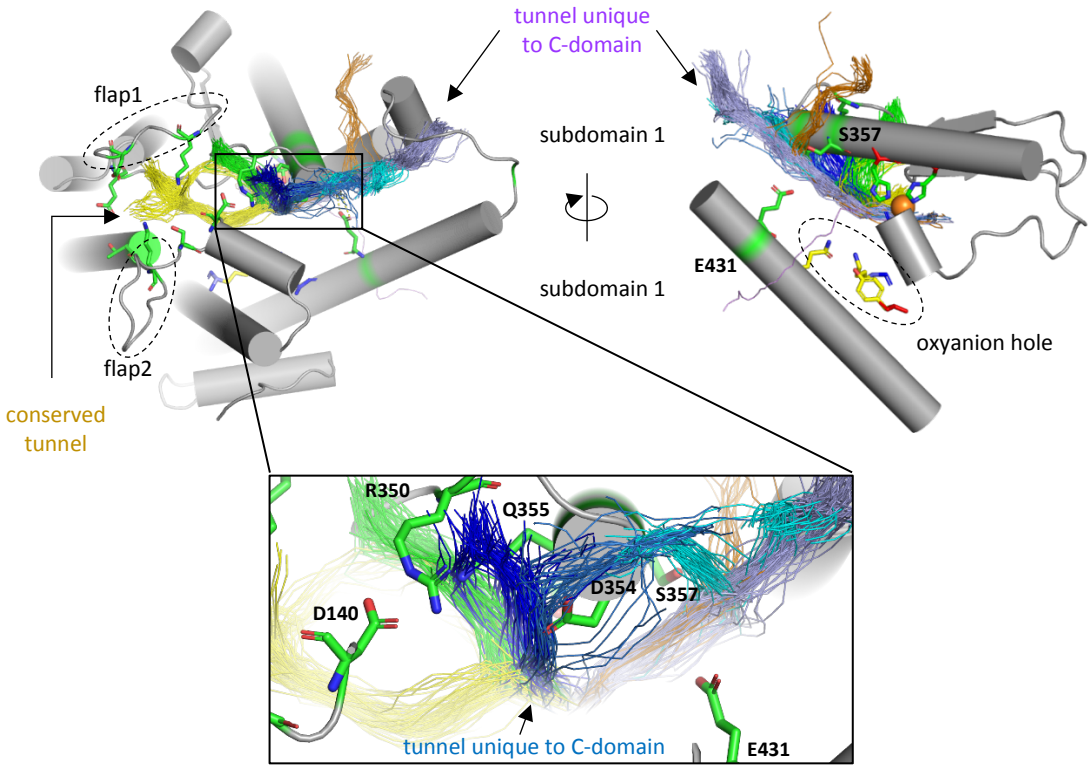


Figure 4.26: Solvent access tunnels unique to the C-domain S_2' subsite.
After alignment of the two domains, N-domain protein (grey cartoon and green sticks) was displayed with C-domain S_2' tunnels obtained from Caver Analyst (lines). All labelled residues are unique to the N-domain (Q355 overlaps with the blue unique C-domain tunnel). Substrate-binding oxyanion hole residues are shown as yellow sticks and Zn as an orange sphere.

A greater number of distinct tunnel clusters were also observed in the C-domain non-prime subsite (Figure 4.25c-d). This likely resulted from the C-domain's increased lid flexibility and lack of c3 binding to subdomain 2. In the N-domain, tunnel formation is restricted by decreased lid and subdomain 2 flexibility via the ionic zipper, subdomain 2 glycan clustering and protein hydrogen bonding to n4 and n8. In bridging between the subdomains, n4 further provides a physical barrier to solvent access which is not duplicated by the more distal C-domain c3 or c6.

Although these calculations are useful to identify residues gating access to the active site, the tunnel-lining residues' chemical properties are ignored and no information is provided regarding tunnel usage. The AQUA-DUCT tool (Magdziarz *et al.*, 2017) was therefore used to trace the influx of water molecules and quantify tunnel usage in the explicitly solvated proteins during 30ns simulation. This agreed with the Caver Analyst results and showed water influx via the prime, non-prime and flap regions (Figure 4.27).

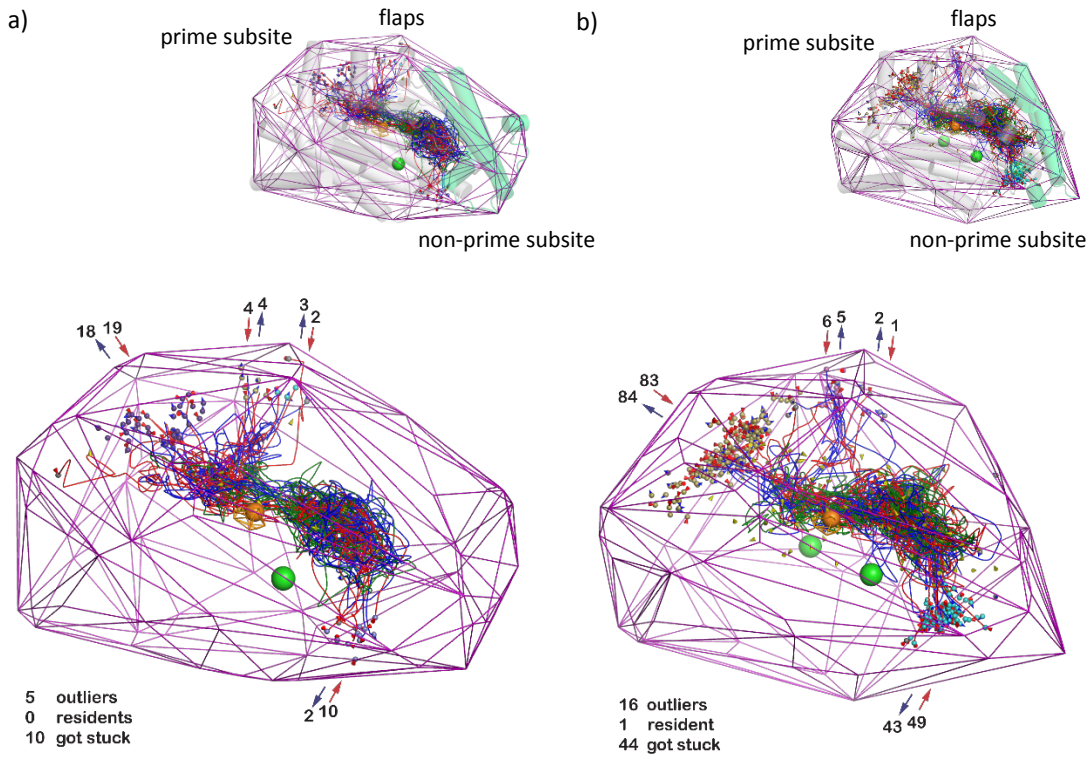


Figure 4.27: Quantitative analysis of water influx during 30ns MD simulation using AQUA-DUCT. a) N-domain and b) C-domain object and scope regions are given as orange and purple convex hulls encompassing the C_α atoms and the Zn-binding site, respectively. Zn and Cl ions are given as orange and green spheres, respectively, and the lid region coloured lime green. Clusters of inlets detected for water molecules during 30ns simulation are displayed as purple, tan, cyan and blue-grey spheres. The averaged smoothed paths followed by these water molecules inside the scope are displayed as lines. Incoming, object and outgoing parts of paths are coloured red, green and blue, respectively. The number of water molecules entering or exiting each tunnel are annotated with arrows. Outliers do not below to any cluster of inlets. The number of water molecules entering the object but never exiting the scope are indicated as well.

The main site for entry and egress in both domains was the prime subsite with very little usage of the flap region tunnels (Appendix A20 Tables A13-A14). This finding is interesting since the ACE crystal structures showed that access to the prime subsite is restricted to a small occluded C-terminal slot while access to the non-prime subsite can occur via a larger

N-terminal pore (Sturrock *et al.*, 2004). Normal mode analyses also suggested that substrate entry occurs via the non-prime subsite through a shift in the N-terminal lid helices' positions (Watermeyer *et al.*, 2006; Anthony *et al.*, 2010). The *in silico* MD results presented here thus suggest for the first time that substrate/solvent entry or product/solvent release can occur through either subsite and that the prime subsite might be preferred. In future, this could have implications for understanding the basis for domain-selective substrate binding since the prime subsite contains numerous unique residues which could gate substrate access.

Water influx to the C-domain active site was increased by 4- to 5-fold at both the prime and non-prime inlets and a significantly greater number of these molecules got stuck (44 never left the scope region again) (Figure 4.27; Appendix A20 Tables A13-A14). They did not reside in distinct pockets and instead roamed widely through the prime and especially the non-prime subsites (Figure 4.28). The space covered by the 10 molecules that got stuck in the N-domain was noticeably narrower. Although the unique N-domain prime subsite residues and glycan n4 do not completely prevent access to the active site, they thus do indeed decrease the influx, as suggested by the Caver Analyst results.

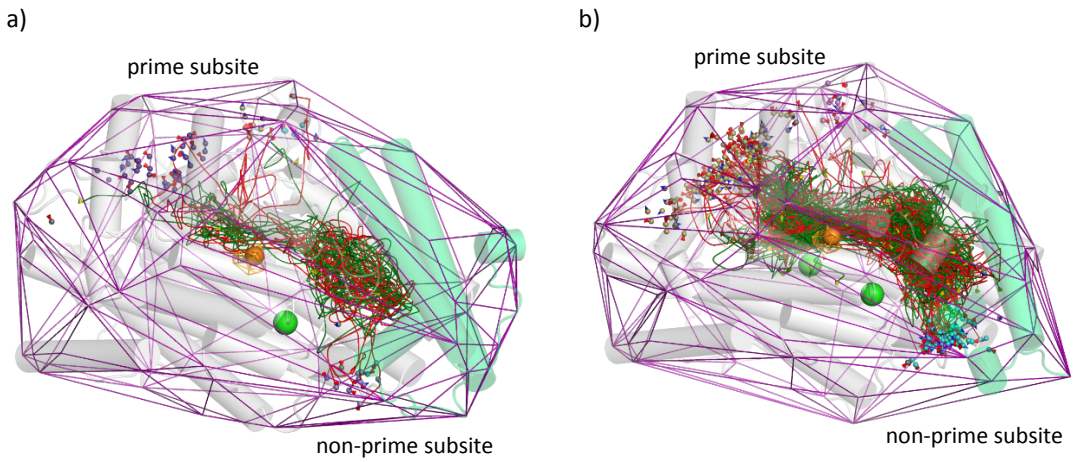


Figure 4.28: Paths followed by water molecules that remain within the scope after entry during 30ns MD simulation (calculated with AQUA-DUCT).
The object and scope regions for a) N-domain and b) C-domain are given as orange and purple convex hulls encompassing the C_α atoms and the Zn-binding site, respectively. Zn and Cl ions are given as orange and green spheres, respectively, and the lid region coloured lime green. Clusters of inlets detected for water molecules during 30ns simulation are displayed as purple, tan, cyan and blue-grey spheres. The smoothed paths followed by water molecules that never exit the scope again are displayed as lines. Incoming and object parts of paths are coloured red and green, respectively.

There is a strong correlation between solvent accessibility and the intrinsic chemical reactivity of an amino acid to oxidative modification (Huang *et al.*, 2015). To propose a mechanism whereby 1,8-ANS binding could protect against oxidative inactivation, the proximity of readily oxidizable active site residues to C-domain tunnels was visualized. The Caver Analyst results were used for simplicity. Values describing the intrinsic reactivity to oxidation (R_i) of each amino acid were taken from Huang *et al* (Huang *et al.*, 2015) and mapped onto the C-domain structure. In the prime subsite, tunnels between the subdomain 1 and 2 flaps and the S_2' subdomain interface could contribute to oxidative inactivation of both domains via oxidation of the substrate-binding oxyanion hole tyrosine (Figure 4.26). The C-domain S_2' tunnels may further increase oxidative inactivation over N-domain by enhancing ROS influx to the readily oxidizable zinc-binding histidine residues (Figure 4.26 and 4.29). Oxidation of these key residues would result in a loss of Zn-coordination and thus, protein inactivation.

Residues in the non-prime subsite were observed to have high reactivities to oxidation (Figure 4.29). Enhanced ROS access via this subsite can also lead to oxidation of the zinc-coordinating histidine residues. In addition to this, however, the non-prime tunnels' proximity to the Cl(II) pocket could result in Y188 and/or R486 oxidation and loss of this critical ACE-activating anion. The high reactivity of these residues suggests that this subsite is the key differentiating factor in the domains' susceptibility to oxidative inactivation. Unique glycans and amino acids endow the N-domain with superior resistance by stabilizing the protein and limiting ROS access.

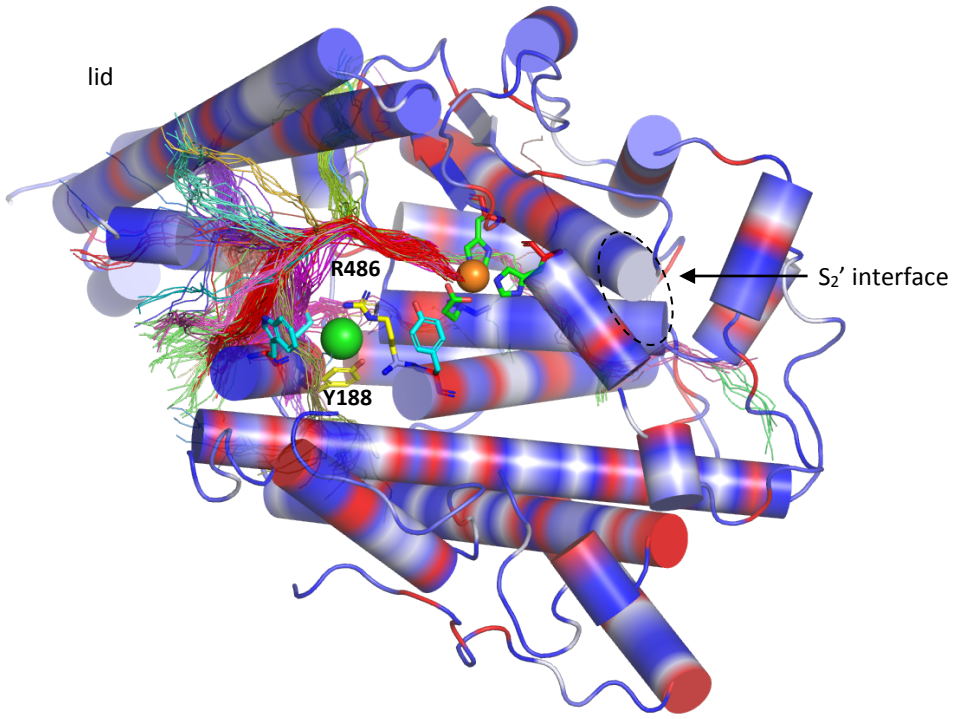


Figure 4.29: Reactivity to oxidation of C-domain non-prime tunnel-lining residues.
C-domain residues are coloured according to their intrinsic reactivity to oxidation (R_i (Huang *et al.*, 2015) where $R_{i\ (blue)} < R_{i\ (white)} < R_{i\ (red)}$) and displayed as a cartoon. The C-domain non-prime tunnels calculated by Caver Analyst are shown as lines, Zn as an orange sphere, Cl^- as a green sphere in pocket II, Zn-coordinating residues as green sticks and Cl^--binding residues as yellow sticks.

Binding of 1,8-ANS to proteins greatly depends upon lysine or arginine salt bridge formation via its sulphonate. This is reinforced by Van der Waals interactions with hydrophobic residues via its anilinonaphthalene (Gasymov and Glasgow, 2007). To understand how 1,8-ANS binding might prevent C-domain oxidative inactivation, the distribution of cationic and aromatic or hydrophobic residues over the protein surface was visualized. Three arginine, two lysine, two tryptophan and one tyrosine residue were found at the C-domain lid's N-terminus, thus 1,8-ANS binding at this site appears likely (Figure 4.30). It is proposed that 1,8-ANS binding to one of these hydrophobic and cationic residues would protect the C-domain against oxidation by stabilizing the lid and limiting solvent access, analogous to N-domain n4. Due to the flexibility of the C-domain lid residues, molecular docking of this molecule was not attempted. In future, however, this hypothesis can be confirmed by C-domain co-crystallization with 1,8-ANS.

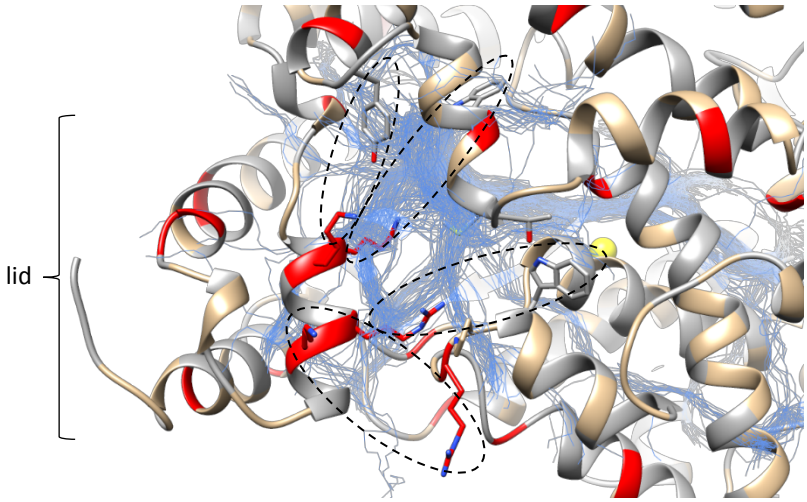


Figure 4.30: Cationic and hydrophobic C-domain residues proximal to the non-prime tunnels. Non-prime tunnels calculated by Caver Analyst are shown as blue lines, hydrophobic residues as grey and cationic residues as red sticks. Zn is shown as a yellow sphere. Potential pairs of 1,8-ANS binding residues are encircled.

4.5 Conclusions

The research presented in this chapter illustrate the power of molecular dynamics simulations in elucidating the function of protein glycosylation. For the first time, atomic-level insight was obtained into the mechanism whereby glycosylation at specific sites contribute to N- and C-domain ACE folding. These *in silico* results were in line with previous experimental work. Additionally, it was found that the N-domain's superior thermal stability and resistance to oxidative inactivation by diffuse radicals is the result of unique protein-protein, glycan-protein and glycan-glycan interactions. This study not only paves the way for determination of the full-length sACE structure but can also aid in the future design of ACE-modulating agents with potential anti-inflammatory activity.

Chapter 5

Concluding remarks and future directions

This study demonstrates how a range of *in vitro* and *in silico* techniques can be integrated to gain insight into the mechanisms of differential ACE folding, thermal stability, inhibition and oxidative inactivation. The molecular basis for N-domain selectivity of two compounds, 33RE and CuGGHLis, from different drug classes was described, with the former resulting in reversible inhibition and the latter irreversible oxidative catalytic inactivation of the N-domain. Additionally, the function of glycosylation at each of the nine N-domain and six C-domain PNGs and the basis for the N-domain's superior thermal stability and resistance to oxidative inactivation by diffuse radicals was elucidated.

This research offered a better understanding of the ACE molecule as a whole and showed that, despite their high active site identity and sequence similarity, the two domains have distinct dynamic motions which could govern enzyme activity and inhibitor binding. The abundance of unique residues at the subdomain interface and hinge regions and the unique locations of the PNGs suggests that, following evolutionary duplication of the *Ace* gene, these regions evolved independently in the two ACE domains. This likely gave rise to their selectivity towards binding or hydrolysis of certain substrates and thus, distinct physiological functions.

The findings of the current study have implications for the future structural elucidation of sACE as well as the design of ACE inhibitors, catalytic metallodrugs and modulating agents with potential inflammatory activity. An outline of the future directions is given in the following subsections.

5.1 Elucidating the molecular basis of N-selective ACE inhibition

Implications for design of domain-selective competitive ACE inhibitors

Using enzyme kinetics, differential scanning fluorimetry, X-ray crystallography and molecular dynamics simulations the molecular mechanism whereby the phosphinic inhibitor 33RE binds with a remarkable 1000-fold selectivity to the N-domain was elucidated. A combination of *in silico* and *in vitro* techniques was required since this drug's binding was mediated by the dynamic behaviour of the N-domain, governed by interactions between distal unique residues. Although the 33RE co-crystal structure of SEDSTE_YR-n389 provided some hints that mutation of these unique residues altered the protein's dynamics, the effect on 33RE binding could only be rationalized using MD simulations. In future, the likelihood of obtaining new domain-selective ACE inhibitors with drug design could be improved by taking the unique dynamic nature of the two domains into account. Usage of weaker zinc binding groups might also allow for increased selection based on interactions of the prime and non-prime moieties with unique residues in these subsites. Given the abundance of unique residues in the S_2' subsite, it is proposed that incorporation of a weak zinc binding group and small, polar P_2' moiety into available moderately N-selective ACE inhibitors might yield clinically-relevant compounds for the treatment of fibrosis.

Design of domain-selective allosteric ACE inhibitors

Domain-selective ACE inhibition might alternatively be achieved by exploiting the lid region's dynamic behaviour. Since hinging of the lid appears to govern inhibitor binding as well as substrate access and hydrolysis, allosteric inhibition of ACE could potentially be achieved by locking the lid in a particular conformation. Recently, nanomolar potency allosteric inhibition of the metalloprotease neurolysin was demonstrated using this approach (Hines *et al.*, 2014). Binding of a pyrazolidine derivative to the lid residues in the protein interior led to non-competitive inhibition by presumably preventing a hinge-like motion following substrate binding. Since ACE possesses a similar hinging motion and sequence conservation between the N- and C-domain is poorest at the lid region, allosteric inhibitor design against the lid appears promising for selective fibrosis and hypertension treatment, respectively.

Alignment of neurolysin co-crystallized with the allosteric inhibitor R2 (PDB ID: 4FXY) to the C-domain of ACE revealed that the 350-fold C-domain selective BPPb (PDB ID: 4APJ) and the 230-fold C-domain selective FII peptide (PDB ID: 2XY9) were bound in the same vicinity as R2 (Figure 5.1). Interestingly, R2 was located next to the lid pivot point identified in Chapter 2 of this thesis. C-domain selective allosteric ACE inhibitors could thus be designed by using FII or the 11-mer BPPb peptide's moieties in this subsite as a template and extending it into the space occupied by R2 to prevent lid tipping about the pivot point. Incorporation of a strong negative charge such as fluorine could further be used to increase the potency of inhibition through displacement of the essential chloride in the second pocket, proximal to the lid. This novel class of ACE inhibitors would not act by chelation of the essential zinc ion and thus should display increased selectivity over other metalloproteases.

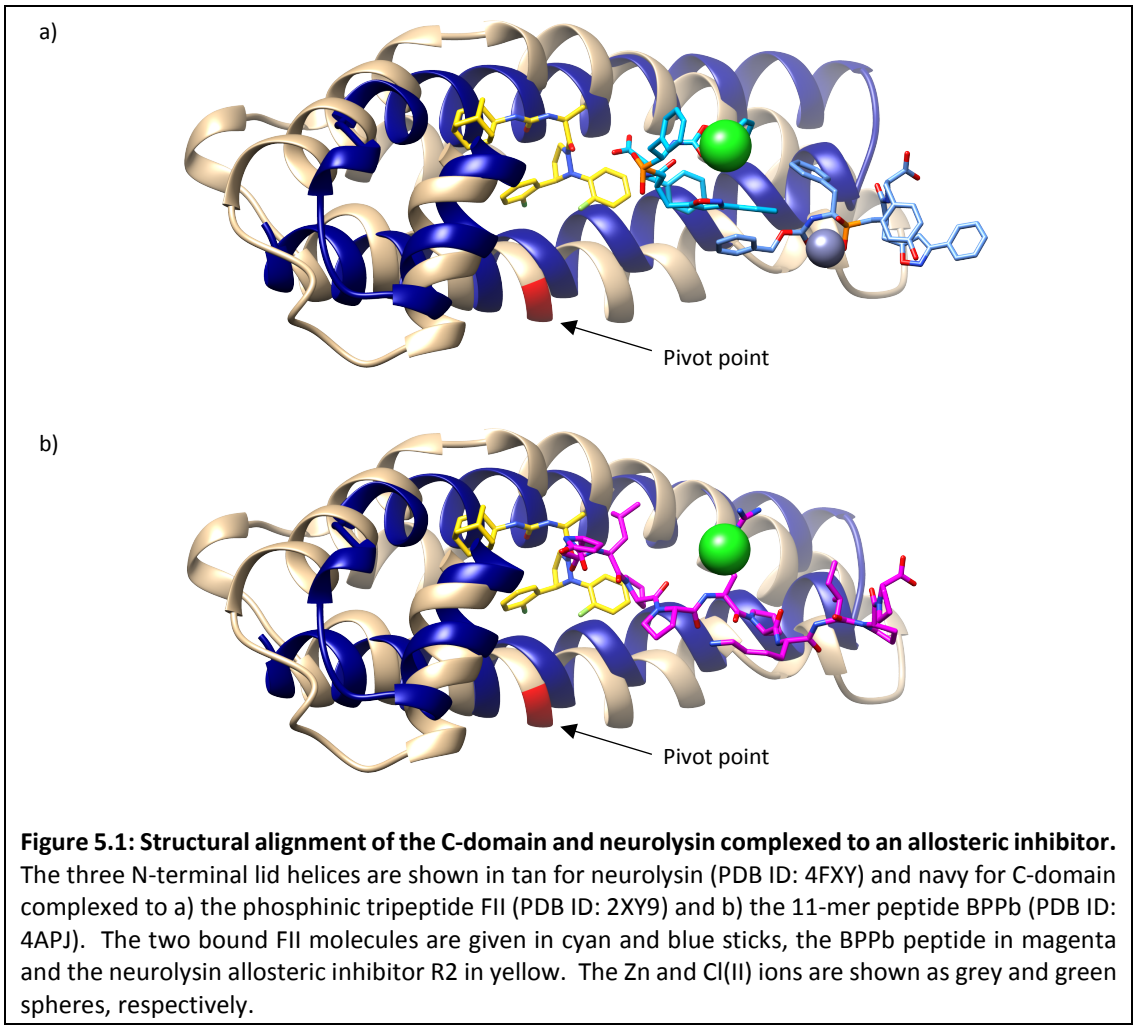


Figure 5.1: Structural alignment of the C-domain and neurolysin complexed to an allosteric inhibitor. The three N-terminal lid helices are shown in tan for neurolysin (PDB ID: 4FXY) and navy for C-domain complexed to a) the phosphinic tripeptide FII (PDB ID: 2XY9) and b) the 11-mer peptide BPPb (PDB ID: 4APJ). The two bound FII molecules are given in cyan and blue sticks, the BPPb peptide in magenta and the neurolysin allosteric inhibitor R2 in yellow. The Zn and Cl(II) ions are shown as grey and green spheres, respectively.

Interestingly, the lid region also displayed the greatest sequence diversity amongst species (Figure 5.2). Recently, Hasan *et al* found that captopril led to the death of *Anopheles gambiae* and *Aedes aegypti* mosquito larvae by inhibition of their larvicidal ACE activity (Abu Hasan *et al.*, 2017). To combat insecticide resistance and control the spread of malaria, it was proposed that existing ACE inhibitors could be modified using rational drug design to selectively target mosquito ACE.

Designing a compound exclusive for binding to the mosquito ACE might, however, prove challenging given the high sequence conservation of the active site amongst species. Since various habitats and organisms would encounter an insecticide, insufficient selectivity could have dire environmental consequences. Captopril and enalaprilat, for example, have been shown to prolong *Caenorhabditis elegans* roundworm and rodent life-span (Kumar *et al.*, 2016; Santos *et al.*, 2009) while captopril, fosinoprilat and lisinopril decreased fertility of the *Crassostrea gigas* pacific oyster (Riviere *et al.*, 2011). The design of allosteric inhibitors against the poorly conserved lid region therefore appears more suitable for the development of insecticides against *Anopheles gambiae* and *Aedes aegypti* mosquito species. In future, this can be achieved by rational drug design following homology modelling of ACE homologues from mosquito, fly, roundworm, rodent, oyster, crab and leech species to the structures of human ACE.

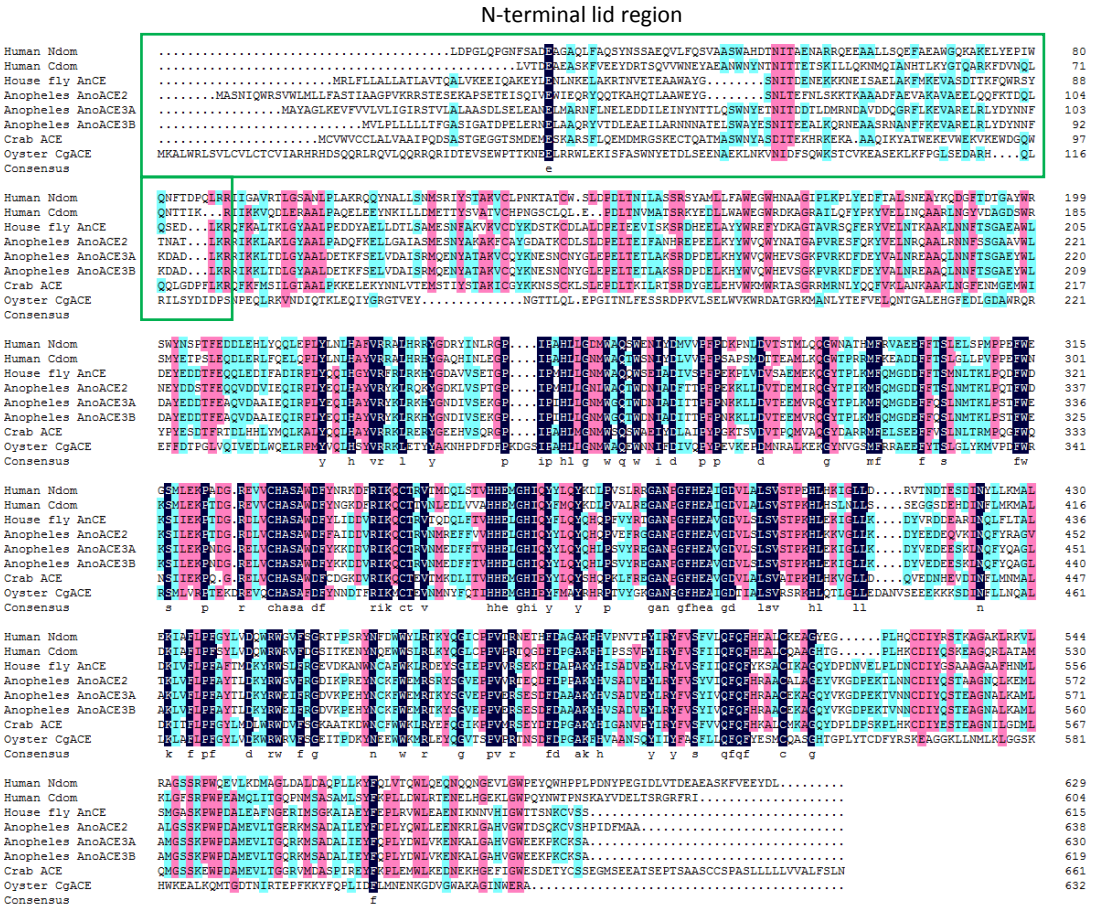


Figure 5.2: Protein sequence alignment between ACE homologues indicating regions of species diversity.
Human ACE homologue sequences from the *Drosophila melanogaster* house fly, *Anopheles gambiae* mosquito, *Carcinus maenas* shore crab, and *Crassostrea gigas* pacific oyster were aligned to the human ACE N- and C-domain using DNAMan v 4.15 (Lynnon BioSoft. ©). The poorly conserved N-terminal lid region is indicated by the green box.

5.2 Evaluating N-selective catalytic inactivation

Identification of the sites of oxidative modifications

The experimental conditions for studying oxidative inactivation of the truncated N- and C-domain ACE proteins by the CuGGHLis metallodrug has been optimized in the current study. The observed concentration-dependent inhibition and time-dependent oxidative inactivation was not confounded by possible inter-domain cooperativity and allowed careful analysis of the inhibitory binding constants and inactivation rate constants. In future, these conditions can be used to prepare inactivated ACE protein samples for mass spectrometric

analysis. This would allow identification of the sites of oxidative amino acid modification and thus further investigations into the molecular basis for N-domain selective inactivation by CuGGHLis. X-ray crystallography can additionally be used to determine the N- and C-domain co-crystal structures with CuGGHLis to visualize the metal-binding domain's orientation to amino acids with high intrinsic reactivities to oxidation.

Design of a C-domain selective catalytic metallodrug

Previously, Hocharoen *et al* observed that upon incubation in the presence of ascorbate and H_2O_2 redox co-reactants, Cu-CB-TE2A-Lis catalysed oxidative inactivation of the N- and C-domain to a similar extent (Hocharoen *et al.*, 2013). This metallodrug bound to both domains with equal affinity and thus also displayed non-selective reversible inhibition. Previously, a lisinopril-tryptophan analogue was designed which was highly selective for inhibition of the C-domain of ACE (Nchinda *et al.*, 2006a). It is proposed that attachment of the CB-TE2A metal-binding domain to the lysyl sidechain of lisinopril-Trp and titration with Cu^{2+} might yield the first C-domain selective catalytic metallodrug for the treatment of hypertension. This compound might be more effective than lisinopril-Trp since it irreversibly inactivates the enzyme and attachment of the hydrophobic chelate could increase the bioavailability of lisinopril-Trp which, at 5.4%, is of potential concern for clinical application (Denti *et al.*, 2014). Time-dependent oxidative inactivation by Cu-CB-TE2A-LisTrp and the mechanism of potential C-domain selectivity can be investigated using the experimental conditions optimized in this study, in addition to mass spectrometry and X-ray crystallography.

5.3 Investigating the influence of *N*-glycans on ACE dynamics

Experimental validation of the proposed glycan functions

In future, the contribution of unique N-domain glycan and protein residues towards protection against thermal and oxidative inactivation can be validated using chimeric constructs. Chimeric C-domain proteins have previously been generated by replacing N-terminal ($C_{1–163}$Ndom-ACE), active site ($C_{164-416}$Ndom-ACE) or C-terminal ($C_{417–579}$Ndom-ACE) sections by their N-domain counterparts (Woodman, 2003; Woodman *et al.*, 2006). Since these areas contained PNGs, this also led to the removal or addition of glycans to the C-domain.

To experimentally validate the contribution of n8 and n9 towards thermal stability and protection against oxidative inactivation, $C_{417–579}$Ndom-ACE can be studied. This chimera retains the C-domain glycans, lid and cleft-lining residues but gains N-domain n8 and n9 (Figure 5.3 a). This would likely introduce a glycan cluster on subdomain 2 through interaction between n9 and the C-domain c4 to stabilize the active site oxyanion hole and protein C-terminus along with n8. Thermal stability would likely be increased but inactivation unchanged since ROS access would still occur via the unique C-domain lid and S_2' tunnels.

The $C_{1–163}$Ndom-ACE construct, on the other hand, can be used to investigate the lid and cleft-lining residues' contribution. Upon replacement of the lid and helix 7 residues with their N-domain counterparts, this chimera lost c1, c2, c3 and c4 but gained N-domain n1, n2, n3, n4, n5 and n6 (Figure 5.3 b). A subdomain 1 cluster would be retained via n2–n3–c5 and a subdomain 2 cluster introduced via n5–n6 interaction. This, along with the addition of n2 and n4 protein interactions, would increase the melting temperature by stabilizing the subdomain 1 flap, lid and subdomain 2. Further stability would be gained by hydrogen bonding between the cleft-lining helices. Apart from this lid stabilization, physical blockade by n4 would also decrease solvent and ROS access through the lid tunnels. Solvent access via the unique C-domain S_2' tunnels, however, would be unchanged.

The unique S_2' tunnels' contribution to oxidative inactivation can be investigated using the fully glycosylated N-domain active site mutant SEDSTE/TSEVVD (Chapter 2). This mutant would maintain N-domain thermal stability and decreased non-prime tunnel access via lid stabilization and n4 blockade. The loss of gating residues upon mutation of the subdomain interface, however, would lead to increased S_2' tunnel formation and an influx of solvent and ROS via this subsite.

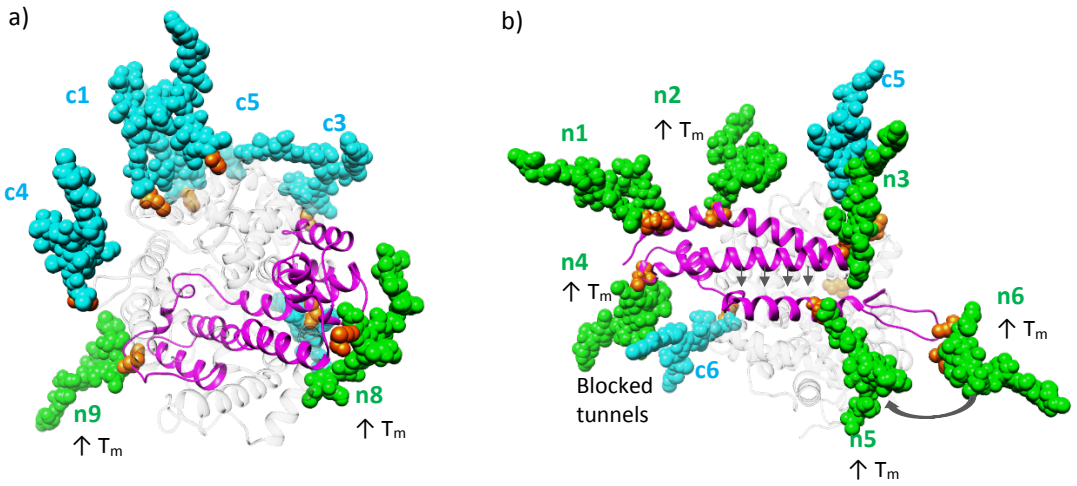


Figure 5.3: Chimeric C-domain proteins for future experimental validation of glycan function. C-domain regions replaced by N-domain residues are shown in magenta, C-domain glycans in cyan and introduced N-domain glycans in green for the a) $C_{417–579}$Ndom-ACE and b) $C_{1–163}$Ndom-ACE chimeric proteins. T_m: C-domain melting temperature. Grey arrows indicate introduced protein or glycan interactions.

Design of anti-inflammatory ACE-modulating agents

Increased susceptibility of the C-domain to oxidative inactivation by diffuse radicals could lead to the progression or maintenance of inflammatory disease (as described in Chapter 4 section 4.1.2). Binding of the 1,8-ANS molecule has been shown to protect the C-domain against ROS and as such could potentially interrupt this inflammatory cycle. Testicular inflammation due to bacterial or viral infection could potentially lead to decreased male fertility via oxidative inactivation of the testis ACE isoform, essentially identical to the C-domain of sACE and known to be important for sperm motility (Hagaman *et al.*, 1998). Although the 1,8-ANS binding site was predicted in the current study, it should be confirmed in future by co-crystallization with the C-domain. Analogues of 1,8-ANS can then be designed using fragment-based drug design or scaffold hopping to obtain selectivity for binding to the C-domain of ACE. This would further require *in silico* screening of the designed compounds against proteins known (from literature) to bind with high affinity to 1,8-ANS. Once a selective compound has been obtained, its ability to protect the C-domain against oxidation by diffuse radicals should be verified *in vitro* using enzyme activity assays. Further *in vitro* experiments can then be performed in cell culture to assess the lead compound's potential for use as an anti-inflammatory agent.

sACE structural elucidation

This study allowed identification of the glycan sites with the largest contribution to N- and C-domain folding and thermal stability. This information can be used in future to guide the design of a minimally glycosylated sACE protein amenable to crystallization for elucidation of its structure. The MD results suggested that a sACE n4,5,7,8,9 c1,3,5 glycoform might be devoid of linker proteolysis and suitable for X-ray crystallography and should thus be created using site-directed mutagenesis.

If it is found, however, that the full glycan complement is required to prevent linker proteolysis, insight can be gained into the human sACE structure by solving the structure of *Anopheles gambiae* AnoACE9. This protein is unique among the insect ACE proteins since it is similar to human sACE in that it consists of two catalytically active domains and possesses a hydrophobic C-terminus (Burnham *et al.*, 2005). It, however, could be more suitable for crystallization since it only contains six PNGs as opposed to the fifteen glycosylated PNGs of sACE. Glycosylation is likely also less complex than in a mammalian system. AnoACE9 shares 38.6% sequence identity and 60.3% sequence similarity with human sACE. Homology modelling of the truncated single-domain ACE structures onto the AnoACE9 structure could therefore be used to gain insight into the possible orientation between the two domains. These two proteins would likely share a similar overall fold as observed between the single domains of human ACE and *Drosophila melanogaster* AnCE (Harrison and Acharya, 2014) (45% sequence identity and 60% sequence similarity to human C-domain ACE).

Elucidation of this structure would have great implications for understanding inter-domain cooperativity, homodimerization, intracellular signalling and, ultimately, the function of sACE in numerous key physiological processes.

Appendices

Chapter 2

A1 Primers used to create active site mutations

Table A1: Details of primers for site-directed mutagenesis.
Screening enzyme recognition sequences are underlined, base-pair changes in lowercase and codon changes in red. T_m: primer melting temperature.

Desired mutant	Template used	Mutation added	Primer sequence (5'-3')	T_m (°C)	Screening enzyme
SE	N-domain	S357V	CGATGGACCAGCTg**gtC**ACAGTGCACCATGAGATGG	78.3	*Pvu*II
	S357V	E431D	GCTAAAAATGGCACTG**GAc**AAAATTGCaTTCCTGCC	75.4	*Bsm*I
SEDSTE	S357V	T358V	CGATGGACCAGCTcGTC**gtA**GTGCACCATGAGATGG	81.1	*Pvu*II removed
	S357V T358V	D354E	CGGGTCACGATG**GAg**CAGCTgGTCGTAGTGCACC	82.3	*Pvu*II
	D354E S357V T358V	E431D	GCTAAAAATGGCACTG**GAc**AAAATTGCaTTCCTGCC	75.4	*Bsm*I
	D354E S357V T358V E431D	E262S	GCCCAGAGCTGG**tcA**AAtATtTACGACATGGTGGTGCC	74.8	*Ssp*I
	E262S D354E S357V T358V E431D	S260T	CATCTGCTGGGAGATATGTGGGCACAG**AcC**TGGTCAAATATTTAC	79.0	*Csi*I
SEDTE	S357V E262S D354E S357V T358V E431D	V357S	GCACACGGGTCAC**c**ATGGAGCAGCTG**tcC**GTAGTGCACC	82.8	*Bst*EII
SEDSTE_R	S260T E262S D354E S357V T358V E431D	R381E	GGATCTGCCTGTaTCCTTGCGT**GAg**GGGGCCAACC	76.6	*Bfu*I
SEDSTE_YR	S260T E262S D354E S357V T358V E431D R381E	Y369F	CCATATACAGTAC**TtC**CTGCAATACAAaGATCTGCCCG	74.2	*Bgl*II

A2 Medium for bacterial growth

Luria Bertani (LB) Medium:

Medium for bacterial growth was prepared by addition of 2.5g tryptone, 1.25g yeast extract, 2.5g NaCl and distilled water to a final volume of 250ml. Medium was autoclaved at 120°C for 20 minutes.

Luria Agar (LA) Plates:

Agar plates for growth of bacterial colonies was prepared by addition of 3.75g agar to 250ml sterilized LB medium. LB-agar was subsequently autoclaved at 120°C (20 minutes) and ampicillin added to a final concentration of 100µg/ml once the solution had cooled to room temperature. Plates were poured and stored at 4°C.

A3 Preparation and transformation of chemically competent cells

Chemically competent bacterial cells:

One hundred microliters of DH5α bacterial cells (from frozen glycerol stock) were inoculated into 5ml LB medium and incubated overnight at 37°C while shaking. The overnight culture was used to inoculate 50ml LB medium which was incubated at 37°C while shaking. The optical density was monitored periodically and the cells removed from the incubator once the log phase of growth was reached (culture optical density of 0.4 – 0.6 at a wavelength of 600nm). The cells were pelleted from culture medium by centrifugation for 10 minutes at a speed of 5000rpm and temperature of 4°C. The pellet was kept on ice for 25 minutes while resuspending the cells in 10ml of 100mM $MgCl_2$. This was followed by centrifugation at a speed of 5000rpm and temperature of 4°C for 10 minutes. After discarding the supernatant, the cells were resuspended on ice in 2ml of a $CaCl_2$ solution consisting of 100mM $CaCl_2$ and 15% glycerol. The resulting competent cells were aliquoted into sterile Eppendorf tubes in 100µl volumes and stored at -80°C.

Bacterial transformation:

Chemically competent DH5α cells were thawed on ice before incubation for twenty minutes on ice with 10µl of plasmid DNA. This was followed by a five-minute incubation at 42°C, incubation on ice for one minute, addition of 450µl LB medium and incubation at 37°C for one hour. One hundred microliters of the resulting transformed cells was spread onto Luria agar plates containing ampicillin and incubated overnight at 37°C.

A4 Plasmid DNA Isolation

Minilysate prep (STET Prep):

A modified version of the minilysate prep by Ausubel *et al* was used to crudely isolate plasmid DNA following cloning or mutagenesis (Ausubel *et al.*, 1989). Briefly, ampicillin-resistant colonies were inoculated into 5ml LB medium (containing ampicillin) and incubated overnight at 37°C while shaking. One millilitre of culture medium was transferred to a sterile Eppendorf tube and centrifuged for two minutes at a speed of 11500rpm. After discarding the supernatant, the pellet was resuspended in 250µl of lysozyme-containing STET buffer (recipe below), briefly vortexed, boiled for one minute and centrifuged for eight minutes at a speed of 11500rpm. The pellet containing cell debris was discarded and plasmid DNA precipitated by addition of 250µl isopropanol. After centrifugation for eight minutes at a speed of 11500rpm, the pellet of extracted plasmid DNA was air-dried and resuspended in 20µl of nuclease-free water.

STET buffer

The buffer used for the abovementioned boiling minilysate prep was prepared by addition of 8g sucrose, 5g Triton X-100, 10ml EDTA (0.5M, pH8.0), 5ml Tris (1M, pH 8.0) and distilled water to a final volume of 100ml.

A5 Agarose gel electrophoresis

A 10X stock of TBE buffer for use in agarose gel electrophoresis was prepared by addition of 53.9g Tris, 27.5g Boric acid, 3.7g EDTA and distilled water to a final volume of 500ml after adjustment of the pH to 8.0. A working solution of 1X TBE buffer containing ethidium bromide (EtBr) was prepared by ten-fold dilution with distilled water and addition of 60µl

EtBr (at 10mg/ml) per litre. A 1% agarose gel was prepared by addition of 0.5g agarose to 50ml of 1X TBE containing EtBr. DNA samples were electrophoresed using 1X TBE buffer containing EtBr as running buffer at a voltage of 80V.

A6 Mammalian cell culture

Tissue culture media were prepared in a fume hood, filter-sterilized and stored at 4°C.

Growth medium

CHO-K1 cell growth medium was prepared by addition of 430ml Dulbecco's Modified Eagle Medium (Gibco®, Invitrogen), 430ml F-12 Ham's nutrient mixture (Gibco®, Invitrogen), 100ml 10%FCS (foetal calf serum inactivated at 56°C for 30min, Sigma-Aldrich® Co.) and 20ml HEPES (1M, pH 7.5).

Harvesting medium

Medium used for CHO-K1 cell protein expression consisted of 475ml Dulbecco's Modified Eagle Medium (Gibco®, Invitrogen), 475ml F-12 Ham's nutrient mixture (Gibco®, Invitrogen), 20ml 10% FCS (foetal calf serum inactivated at 56°C for 30min, Sigma-Aldrich® Co.), 20ml HEPES (1M, pH 7.5) and 10ml Pen-Strep (penicillin-streptomycin, Lonza Research Solutions).

Wash solution

CHO-K1 cells were washed during culture using a phosphate buffered saline solution prepared by addition of 8.0g NaCl, 0.20g KCl, 1.78g Na_2HPO_4, 1.36g KH_2PO_4 and distilled water to a final volume of 500ml after pH adjustment to 7.5.

A7 His-Leu standard curve in phosphate assay buffer

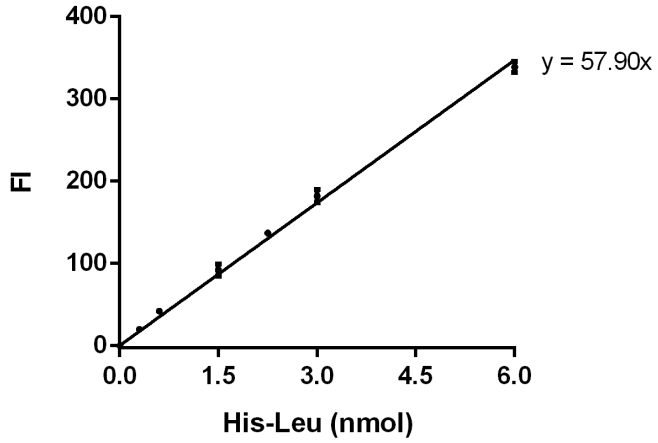


Figure A1: Standard curve of the linear correlation between moles of His-Leu and fluorescence determined in phosphate assay buffer pH 8.3.
Represented as mean ± SEM of triplicate readings (Analysed with GraphPad Prism® v 6.01).

A8 Sodium dodecyl sulphate - polyacrylamide gel electrophoresis

Proteins were separated according to molecular weight using sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE). Sample buffer consisted of 0.0625M Tris-HCl (pH 6.8), 2% SDS, 10% Glycerol, 5% Mercaptoethanol and 0.001% Bromophenol Blue. Five microliters of sample buffer was added per protein sample. The protein was denatured by boiling for 5 minutes.

A 10% running gel was prepared by addition of 3.3ml running gel buffer (consisting of 1.125M Tris-HCl (pH 8.8) and 0.3% SDS), 4.07ml distilled water, 2.5ml acrylamide (at 40%), 100µl of ammonium persulphate (AMPS at 0.1%) and 15µl *N,N,N',N'*-tetramethylethylenediamine (TEMED). A 3% stacking gel was prepared by addition of 3.33ml stacking gel buffer (consisting of 0.375M Tris-HCl (pH 6.8) and 0.3% SDS), 5.2ml distilled water, 1.1ml acrylamide, 300µl of AMPS (at 0.1%) and 20µl TEMED. The gel was cast, samples loaded and electrophoresis performed in a buffer consisting of 0.025M Tris-HCl (pH 8.3), 0.192M glycine and 0.1% SDS at 25mA.

A Coomassie solution consisting of 0.25% Coomassie Brilliant Blue, 50% methanol and 10% acetic acid was used to stain the gel for one hour while shaking. It was subsequently de-stained by incubation in a solution of 25% ethanol and 10% acetic acid for two hours while shaking.

A9 Abz-Gly standard curve

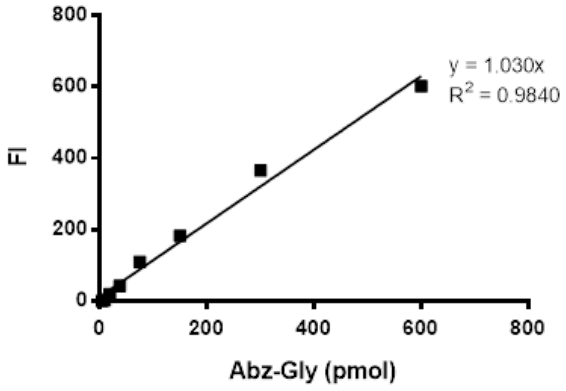


Figure A2: Standard curve of linear correlation between moles of Abz-Gly and fluorescence.
Represented as mean ± SEM of triplicate readings (Analysed with GraphPad Prism® v 6.01).

A10 Abz-Gly correction curve

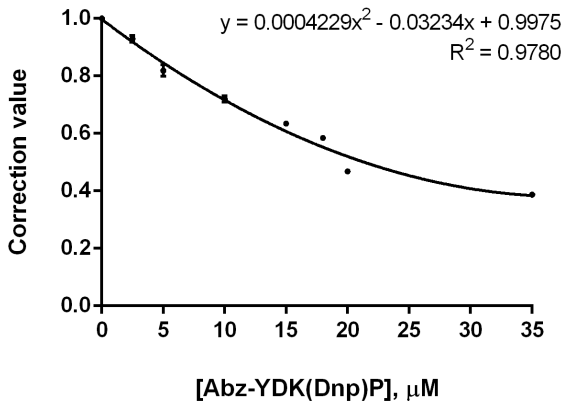
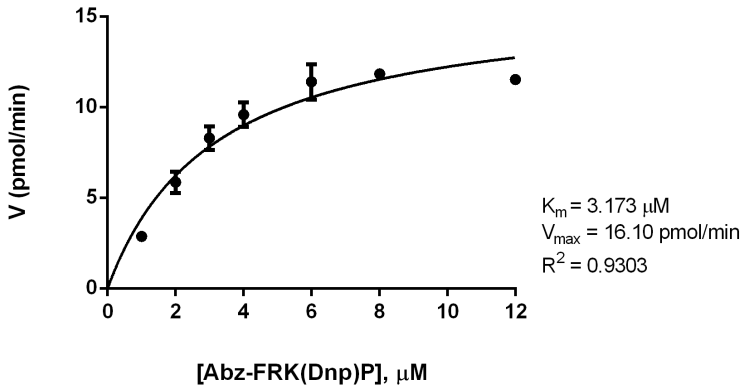


Figure A3: Inner filter effect correction curve.
Correction values represent the ratio of Abz fluorescence at increasing substrate concentration relative to that in the absence of substrate. The data was fit to a second order polynomial and is represented as mean ± SEM from triplicate readings (analysed using GraphPad Prism® v 6.01).

A11 Example of (Abz)-FRK(Dnp)P-OH kinetic characterization

a)



b)

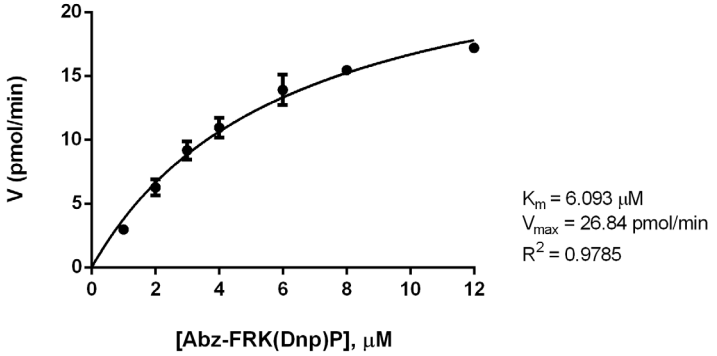


Figure A4: Representative Michaelis-Menten graph illustrating Abz-FRK(Dnp)P K_m determination. N-domain a) Initial uncorrected data and b) data corrected for the inner filter effect are represented as mean ± SEM from triplicate wells (GraphPad Prism® v 6.01 analysis).

A12 Example of a Dixon plot used in characterizing inhibitor binding affinity

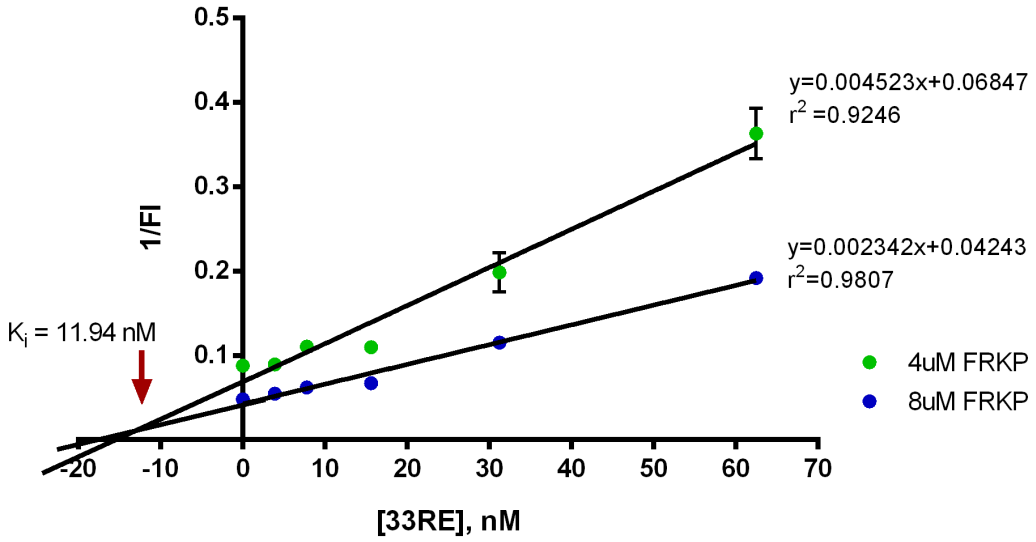


Figure A5: Representative Dixon plot of N-domain titrated with 33RE, illustrating K_i determination. The inverse of initial reaction velocity is represented as 1/FI, where FI = fluorescence intensity. Extrapolation of linear trendlines from different substrate concentrations gives a point of intersection (indicated by the red arrow) representing $-K_i$ on the x-axis, thus K_i = 11.94nM. Data represented as mean ± SEM from triplicate readings (GraphPad Prism® v 6.01 analysis).

Chapter 3

A13 Titrations of EDTA and GGHLis with Cu^{2+}

Standardization of a $CuCl_2$ solution

To obtain a copper solution of known concentration for titration of Gly-Gly-His-Lisinopril (GGHLis), volumetric standardization was employed. An EDTA solution of known concentration was titrated with $CuCl_2$ and the formation of the CuEDTA complex monitored spectrophotometrically. Upon addition of Cu^{2+}, an increase in absorbance was detected at a wavelength of 270nm (Figure A6). After 65µl of $CuCl_2$ was added (pink line on Figure A6), a slight shift in the baseline was observed and further titration yielded no additional increase in the 270nm peak. This indicated that all EDTA molecules were complexed to copper.

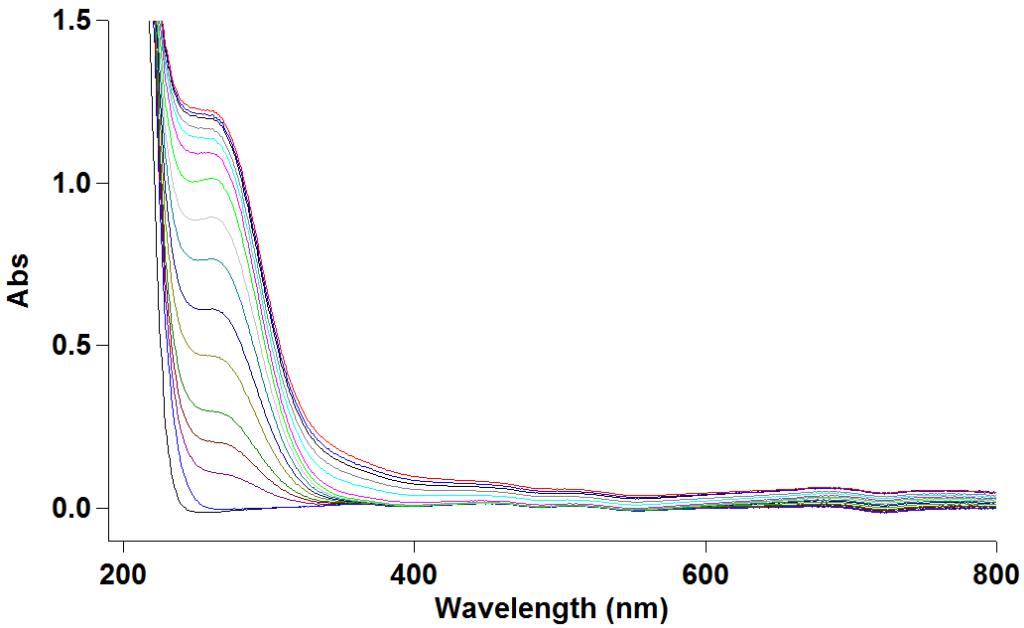


Figure A6: Titration of an EDTA solution with $CuCl_2$.
Absorbance (Abs) of buffer only is given in black and EDTA only in blue. Absorbance after sequential addition of 10µl volumes of $CuCl_2$ is given in purple to red. Titration was performed in buffer consisting of 20mM HEPES pH 7.5 and 100mM NaCl. An increase in absorbance at 270nm indicates formation of the CuEDTA complex.

The absorbance data from this complete titration was subsequently fitted to a segmental linear regression model relative to the molar ratio of metal to chelator. The titration end-point was indicated by the inflection point on the graph (Figure A7). Since it was below 1.00,

the actual Cu^{2+} concentration in the prepared $CuCl_2$ stock solution was slightly greater than the intended 3mM.

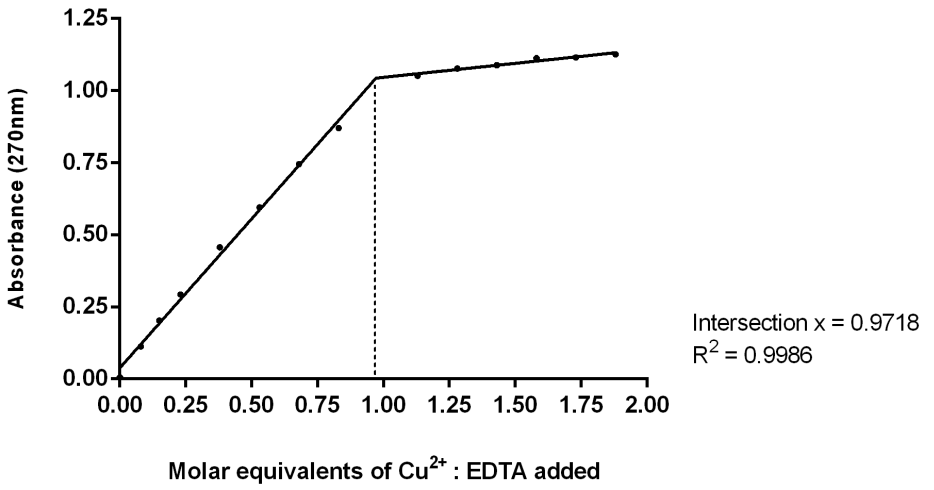


Figure A7: Establishing the titration end-point for the CuEDTA complex.
Complex formation was monitored as an increase in absorbance at 270nm. The data was fitted to a segmental linear regression model using GraphPad Prism® v6.01. The titration end-point was determined from the inflection point and the molar ratio used to calculate the stock solution's actual Cu^{2+} concentration.

Preparation of the CuGGHLis metallodrug

The CuGGHLis metallodrug was prepared by titration with the EDTA-standardized $CuCl_2$ solution. Firstly, the absorbance peak detected at 257nm (Figure A8) was used to confirm the GGHLis concentration using an extinction coefficient of $197M^{-1}cm^{-1}$, due to the presence of the lisinopril phenylalanine moiety. Subsequent addition of $CuCl_2$ into the GGHLis solution led to a shift in the 257nm peak and an increase in absorbance at a wavelength of 250nm (Figure A8). The d-d transition was also detected as an increase in absorbance at 525nm (Lau *et al.*, 1974). The CuGGHLis titration end-point was determined by fitting the absorbance of the more pronounced 250nm peak to a segmental linear regression model (Figure A9). At the inflection point on the curve, and thus the titration end-point, a 1:1 molar ratio of metal to chelate was observed. Further addition of $CuCl_2$ led to a slight shift in the baseline (Figure A8), indicating that all GGHLis molecules were complexed and free Cu^{2+} possibly present in solution past this point.

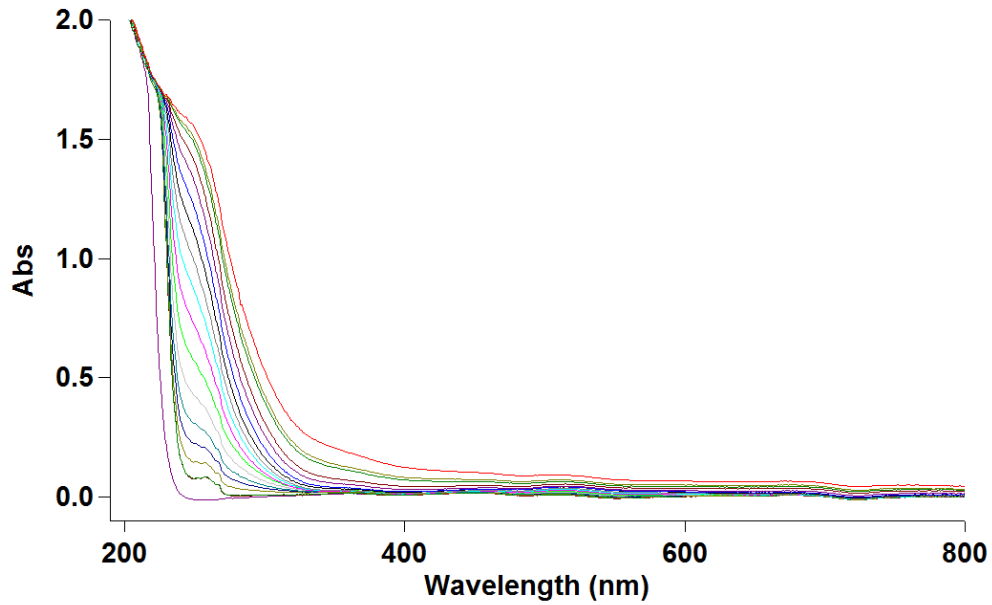


Figure A8: Titration of the GGHLis chelate with $CuCl_2$.
Absorbance (Abs) of buffer only is given in purple and GGHLis only in black. Absorbance after sequential addition of 10µl volumes of $CuCl_2$ is given in brown to red. Titration was performed in buffer consisting of 20mM HEPES pH 7.5 and 100mM NaCl. An increase in absorbance at 250nm indicates formation of the CuGGHLis complex. The d-d transition is evident from an increase in absorbance at 525nm.

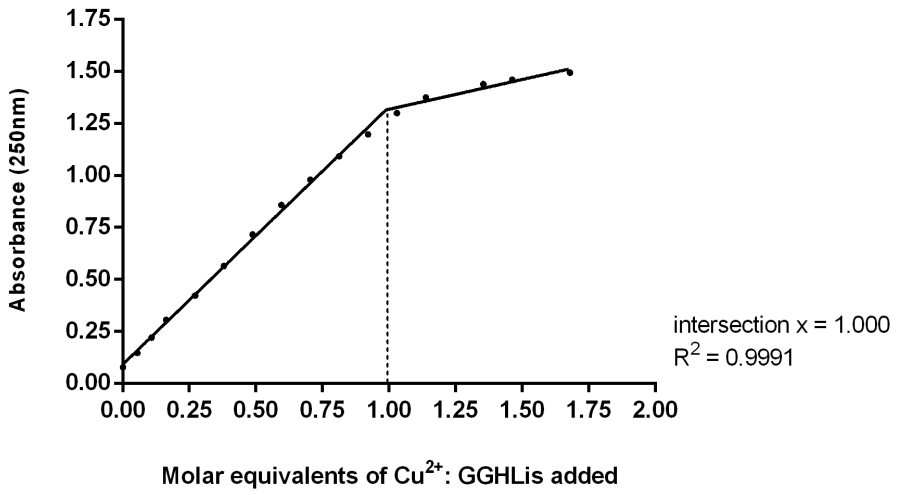


Figure A9: Establishing the titration end-point for the CuGGHLis complex.
Complex formation was monitored as an increase in absorbance at 250nm and the data fitted to a segmental linear regression model using GraphPad Prism® v6.01. The titration end-point was determined from the inflection point.

The second titration was performed up until this established endpoint to obtain a 1:1 molar ratio of metal to chelator (Figure A10) whereas the third titration was halted prior to the endpoint at a 1:1.1 metal to chelator molar ratio (Figure A11). In both cases the lack of free copper ions was confirmed by the absence of absorbance at 750 – 800nm. The third titration yielded a slight excess of chelate to further ensure that no free copper ions would be present during subsequent oxidative inactivation experiments.

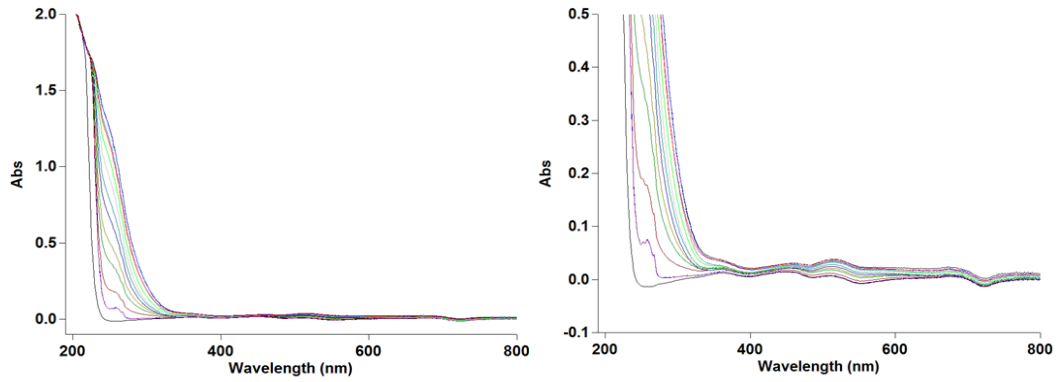


Figure A10: Titration of GGHLis with $CuCl_2$ to prepare the CuGGHLis metallodrug used for binding affinity characterization.
Copper addition was halted at a 1:1 molar ratio of metal to chelator. The zoomed in view in the right-hand panel shows a lack of absorbance at 750-800nm, thus no free Cu^{2+} is present.

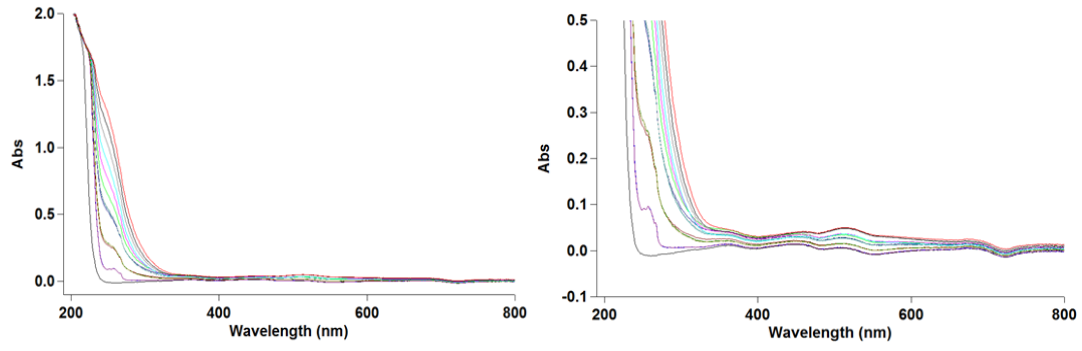


Figure A11: Titration of GGHLis with $CuCl_2$ to prepare the CuGGHLis metallodrug used for characterization of catalytic inactivation.
Copper addition was halted at a 1:1.1 molar ratio of metal to chelator. The zoomed in view in the right-hand panel shows a lack of absorbance at 750-800nm, thus no free Cu^{2+} is present.

A14 His-Leu standard curve in HEPES assay buffer

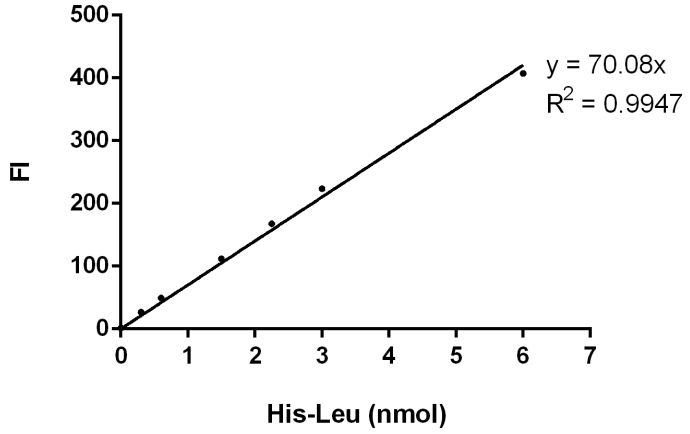


Figure A12: His-Leu standard curve prepared in HEPES pH 8.2 assay buffer (containing Brij at 0.05%). Triplicate readings are shown as mean ± SEM.

A15 Blue Silver Coomassie SDS-PAGE stain

The colloidal Blue-Silver Coomassie stain was prepared according to the method of Candiano *et al* (Candiano *et al.*, 2004). A 200ml volume of stain was prepared by dropwise addition of 20ml *o*-phosphoric acid (Sigma-Aldrich® Co.) to 20ml MilliQ deionized distilled water while stirring. Twenty grams of ammonium sulphate was added while stirring. Once dissolved, 0.24g of Coomassie brilliant blue G250 was added and completely dissolved. The volume was adjusted to 80% of the final volume with MilliQ deionized distilled water before addition of 40ml anhydrous methanol while stirring to obtain colloidal particles in greenish-blue solution. The staining solution was stored in the dark at room temperature.

After separation of the proteins by SDS-PAGE, the gel was washed three times for 10 minutes with MilliQ deionized distilled water to remove SDS. The staining solution was added and incubated at room temperature with shaking to allow distribution of the colloidal particles. To achieve maximum staining, the gel incubations were performed overnight and de-stained by rinsing with MilliQ deionized distilled water.

A16 Reported versus published second order inactivation rate constants

Table A2: Initial rates and second order catalytic inactivation rate constants (k_2) at 37°C for N-domain in the presence or absence of CuGGHLis and redox co-reactants.
Initial rates were calculated from the first-order exponential decay curves obtained from one time-course experiment performed in triplicate and rate constants as k_2 = (initial rate - background)/ [enzyme][CuGGHLis]. A comparison is given with the k_2 values reported by Joyner *et al* at 37°C for two-domain sACE using a non-domain selective substrate.

N-domain		Initial rate (nM min^{-1})		k_2 for inactivation (M^{-1} min^{-1})	
		raw	- background	N-domain	sACE (Joyner *et al.*, 2012)
- CuGGHLis	**Ascorbate**	0.0016			
	H_2O_2	0.0040			
	Ascorbate + H_2O_2	0.0210			
+ CuGGHLis	**None**	0.010	0.010	15 299	
	Ascorbate	0.064	0.062	94 876	102 000 ± 6 000
	H_2O_2	0.014	0.010	14 948	28 000 ± 9 000
	Ascorbate + H_2O_2	0.150	0.130	199 082	152 000 ± 7 000

Chapter 4

A17 System equilibration

A17.1 N-domain energy, temperature, pressure, volume and density

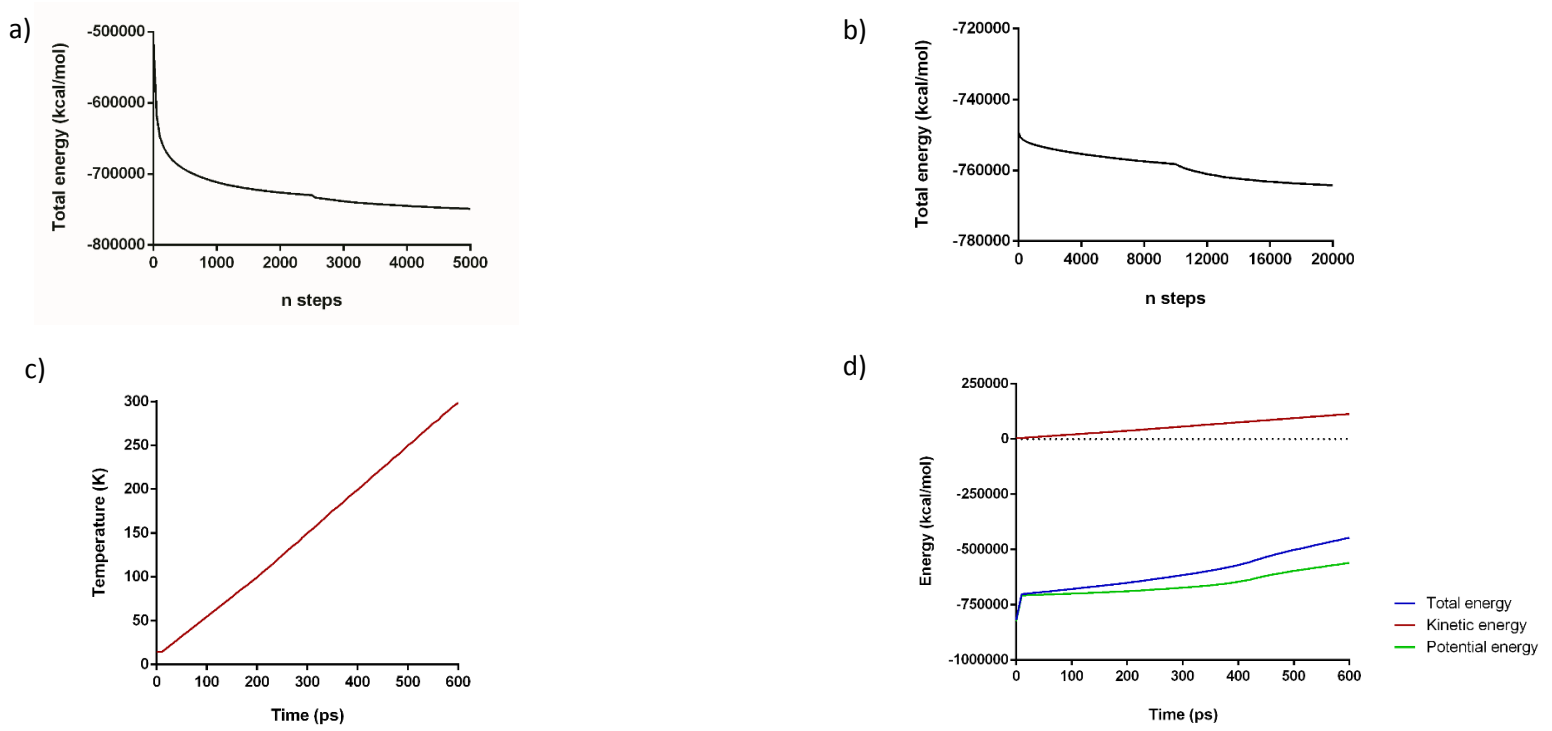


Figure A13: Variation in N-domain energy and temperature observed during (a) minimization 1 (b) minimization 2 and (c-d) heating.

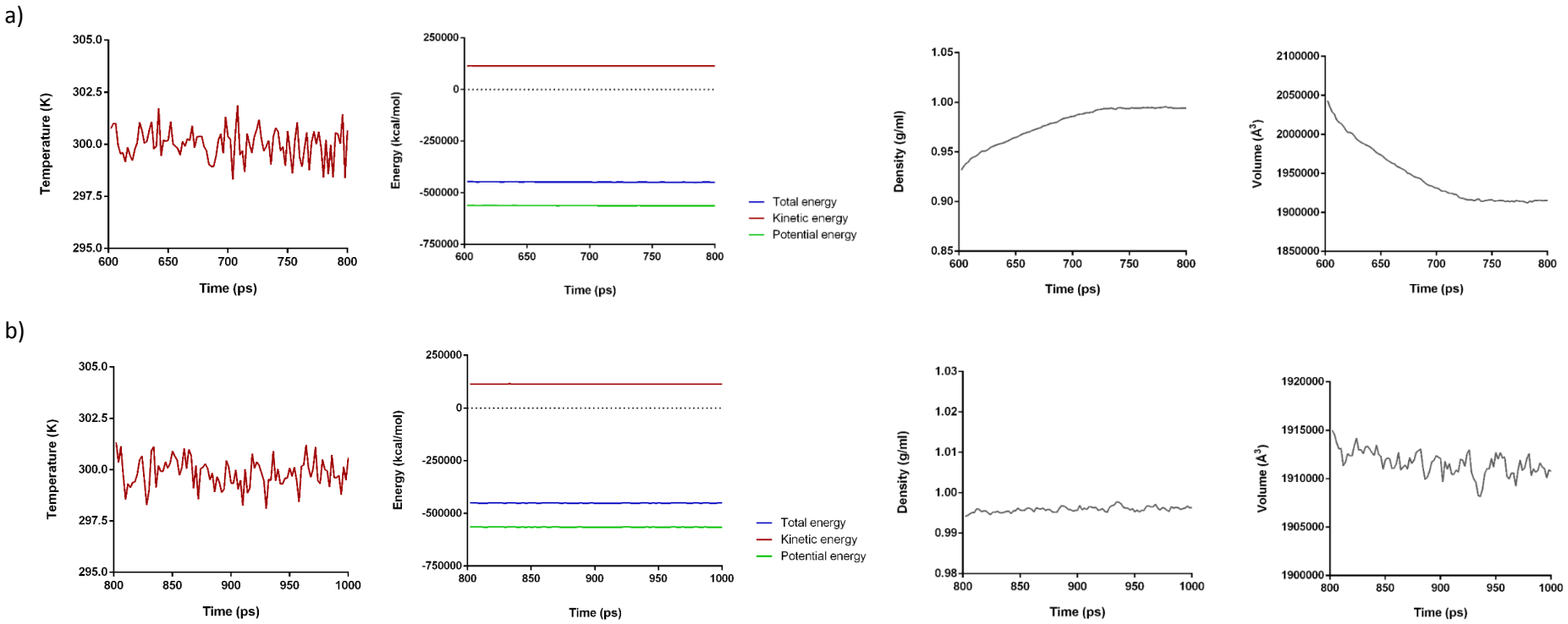


Figure A14: Variation in N-domain temperature, energy, density and volume observed during (a) equilibration 1 (b) equilibration 2.

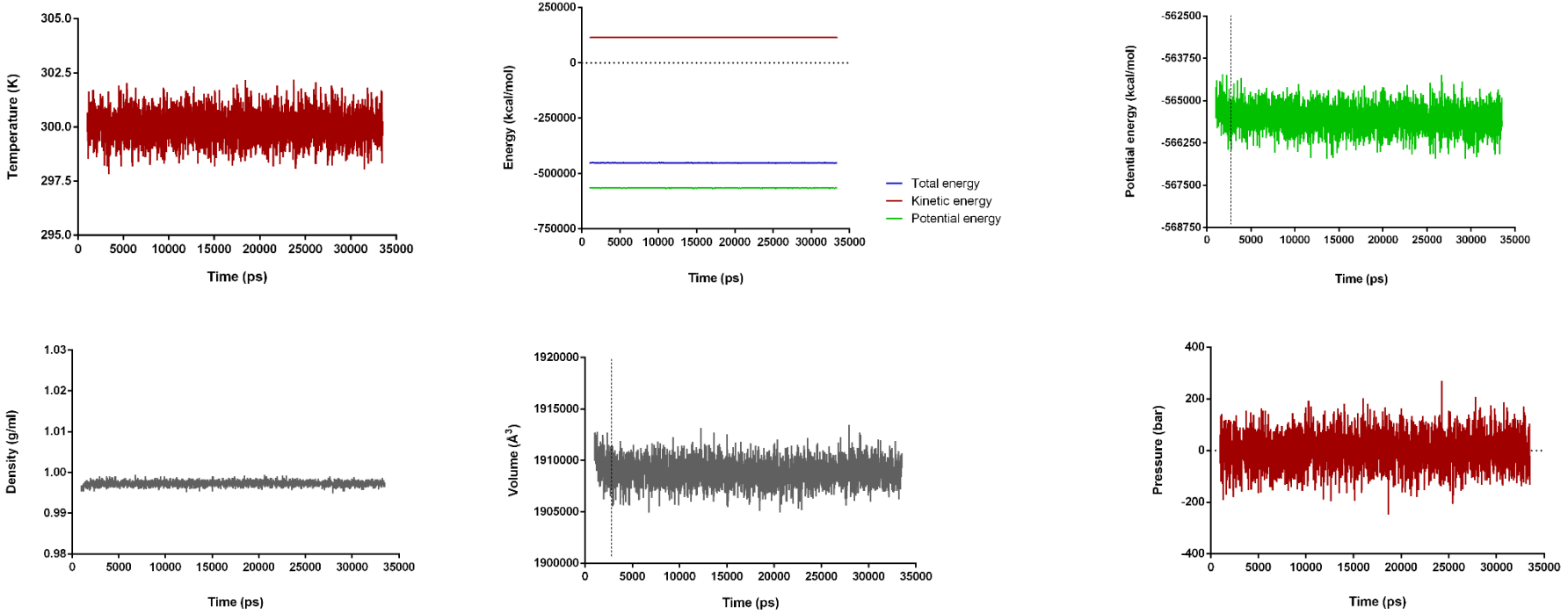


Figure A15: Variation in N-domain temperature, energy, density, volume and pressure observed during 30ns production dynamics.
The vertical dotted line indicates the point where the system was considered equilibrated and from which trajectory analysis was performed.

A17.2 C-domain energy, temperature, pressure, volume and density

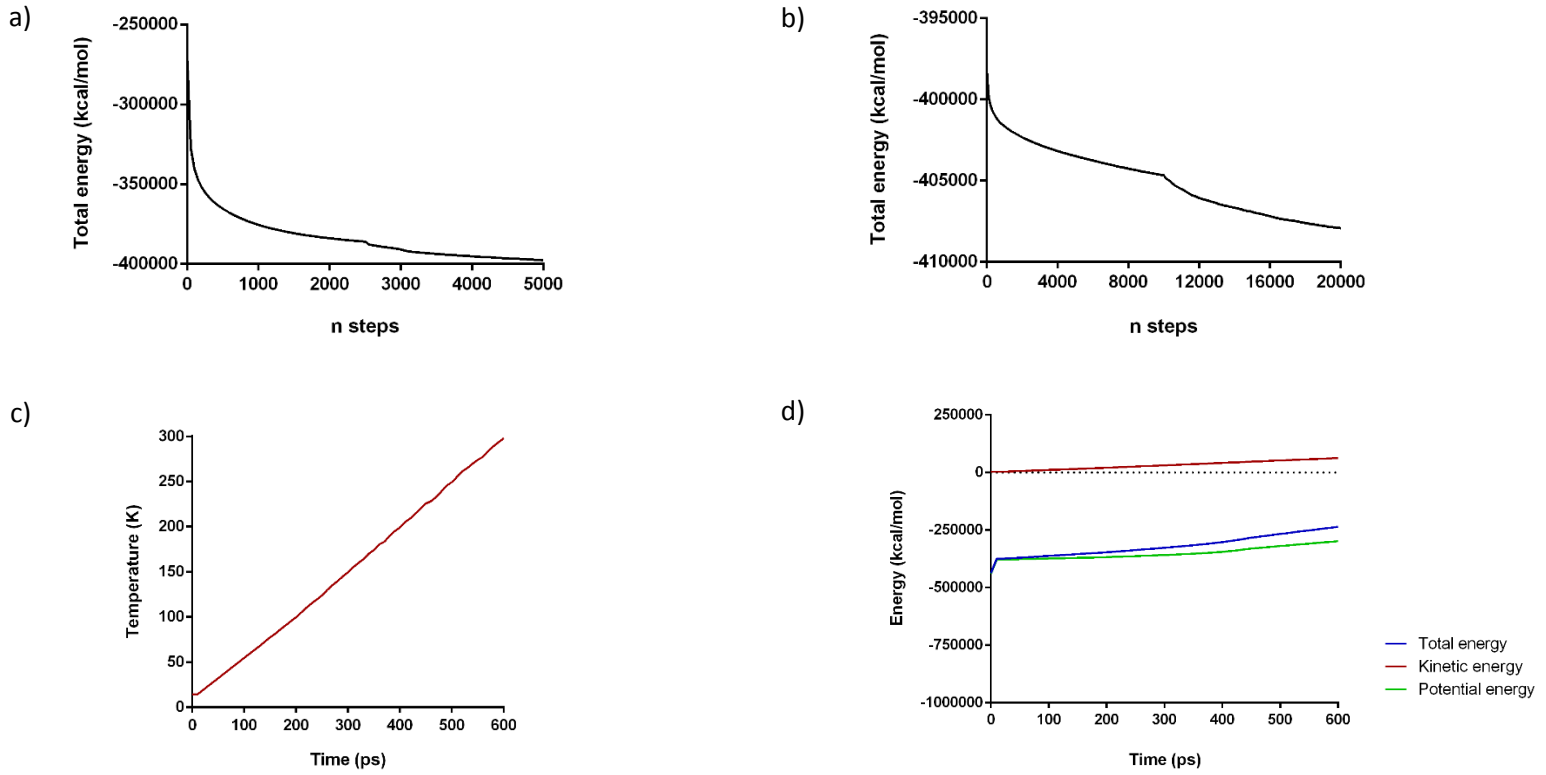


Figure A16: Variation in C-domain energy and temperature observed during (a) minimization 1 (b) minimization 2 and (c-d) heating.

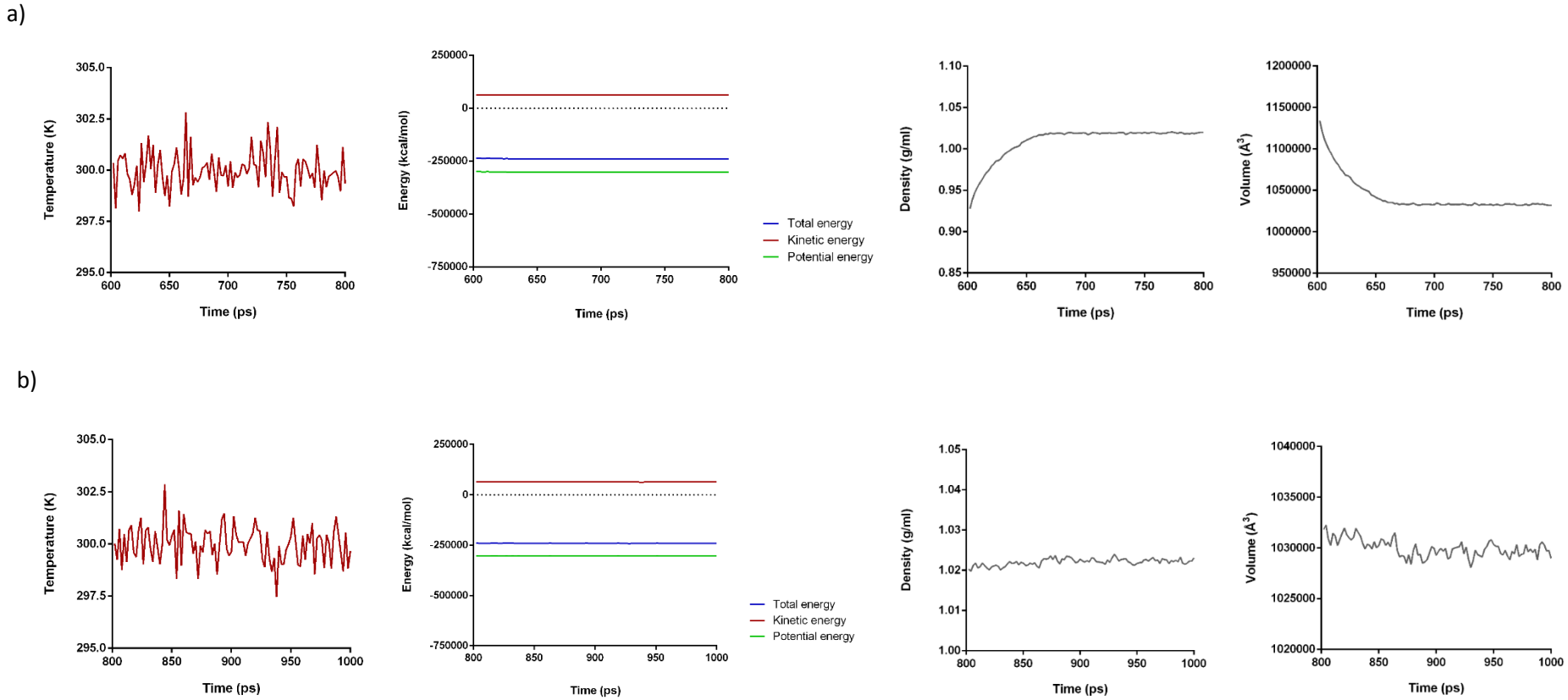


Figure A17: Variation in C-domain temperature, energy, density and volume observed during (a) equilibration 1 (b) equilibration 2.

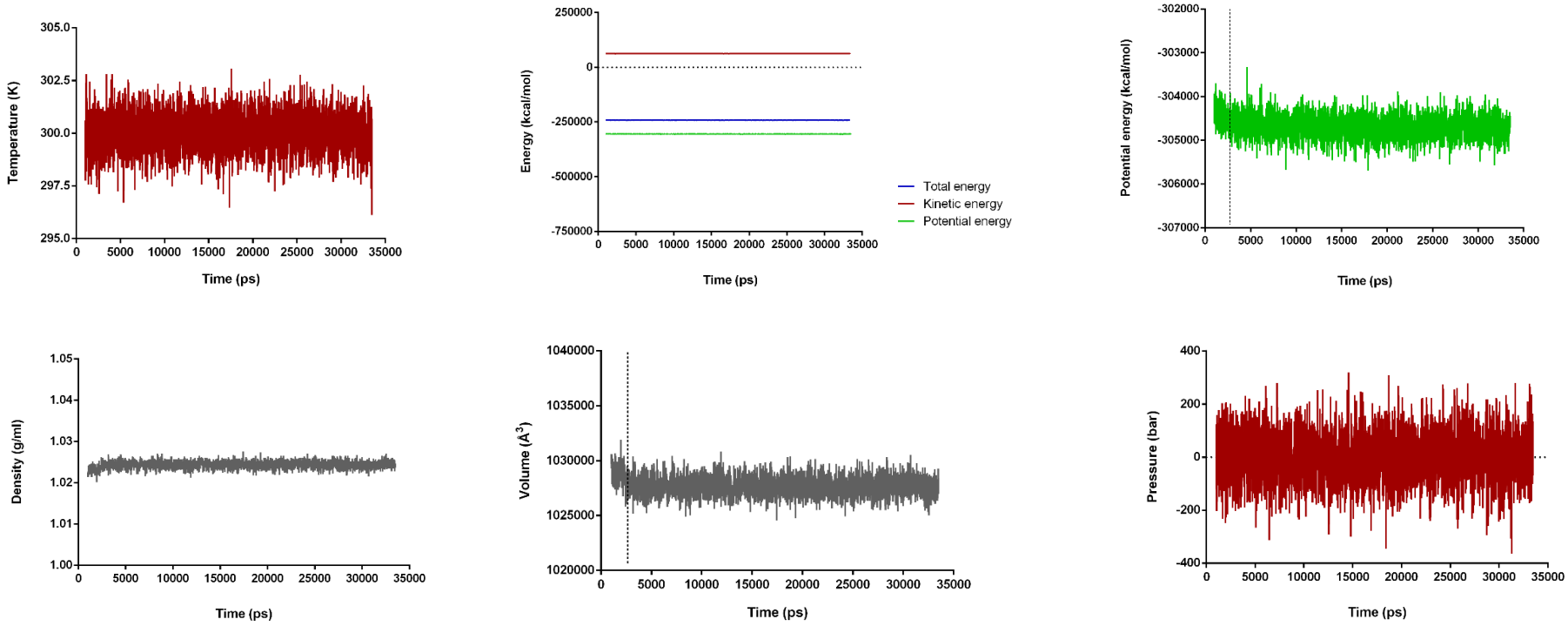


Figure A18: Variation in C-domain temperature, energy, density, volume and pressure observed during 30ns production dynamics.
The vertical dotted line indicates the point where the system was considered equilibrated and from which trajectory analysis was performed.

A17.3 Maintenance of N-domain Zn-coordination geometry

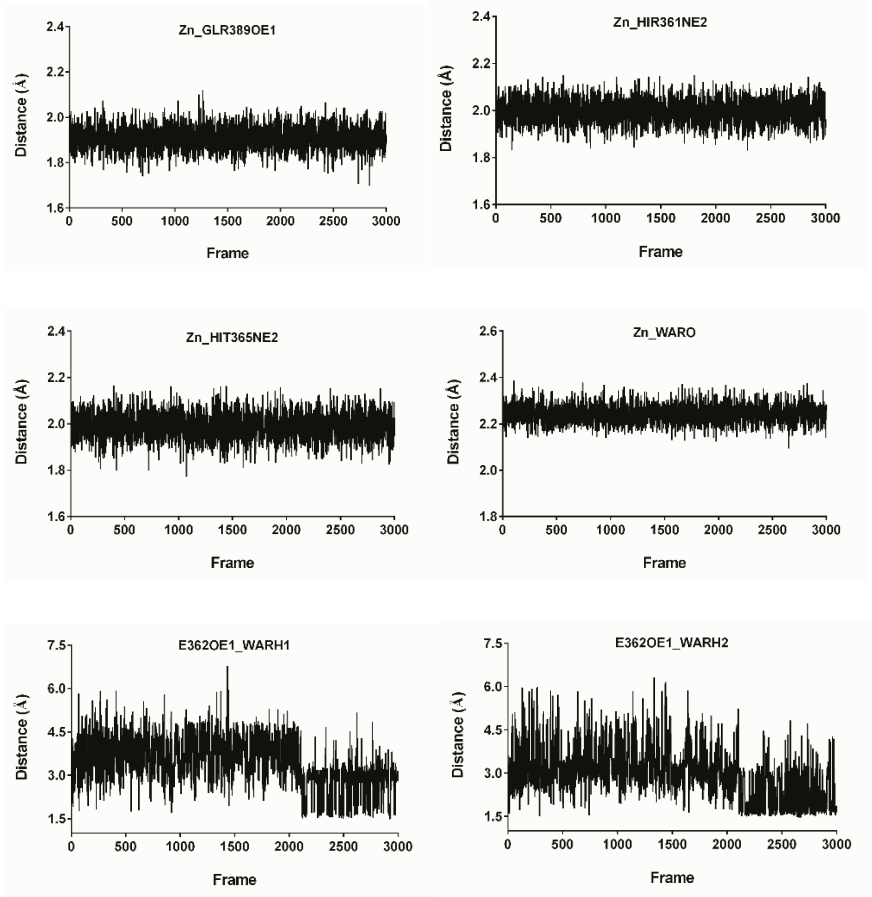


Figure A19: Variation in N-domain Zn-coordinating bond distances during 30ns simulation.

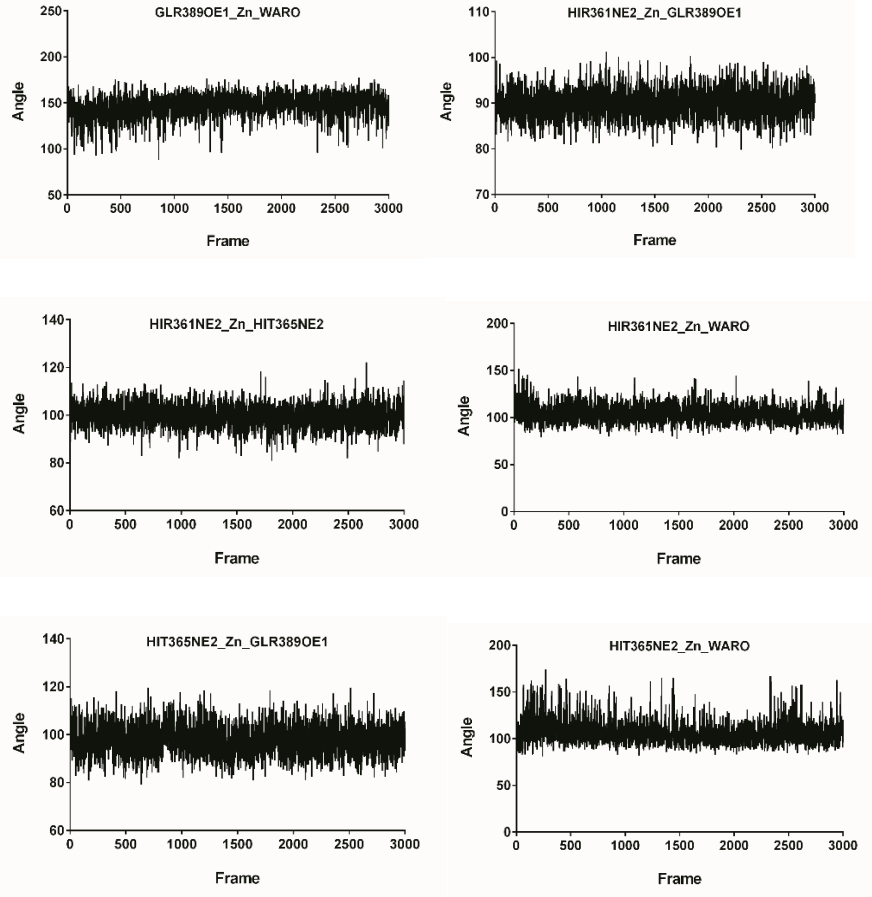


Figure A20: Variation in N-domain Zn-coordinating bond angles during 30ns simulation.

A17.4 Maintenance of C-domain Zn-coordination geometry

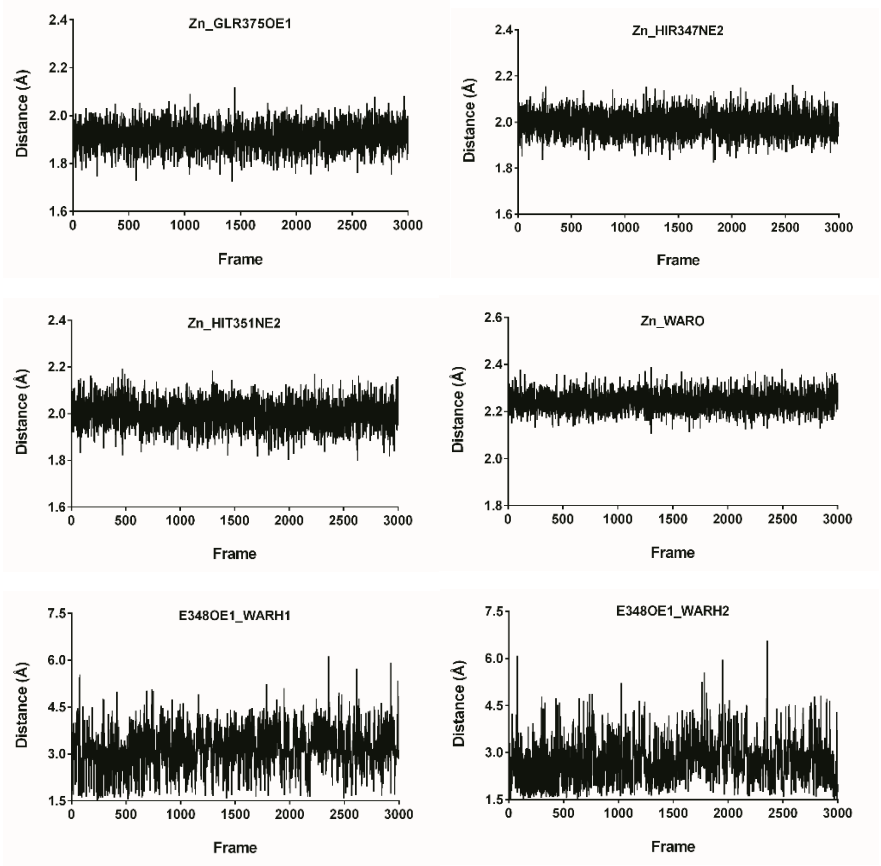


Figure A21: Variation in C-domain Zn-coordinating bond distances during 30ns simulation.

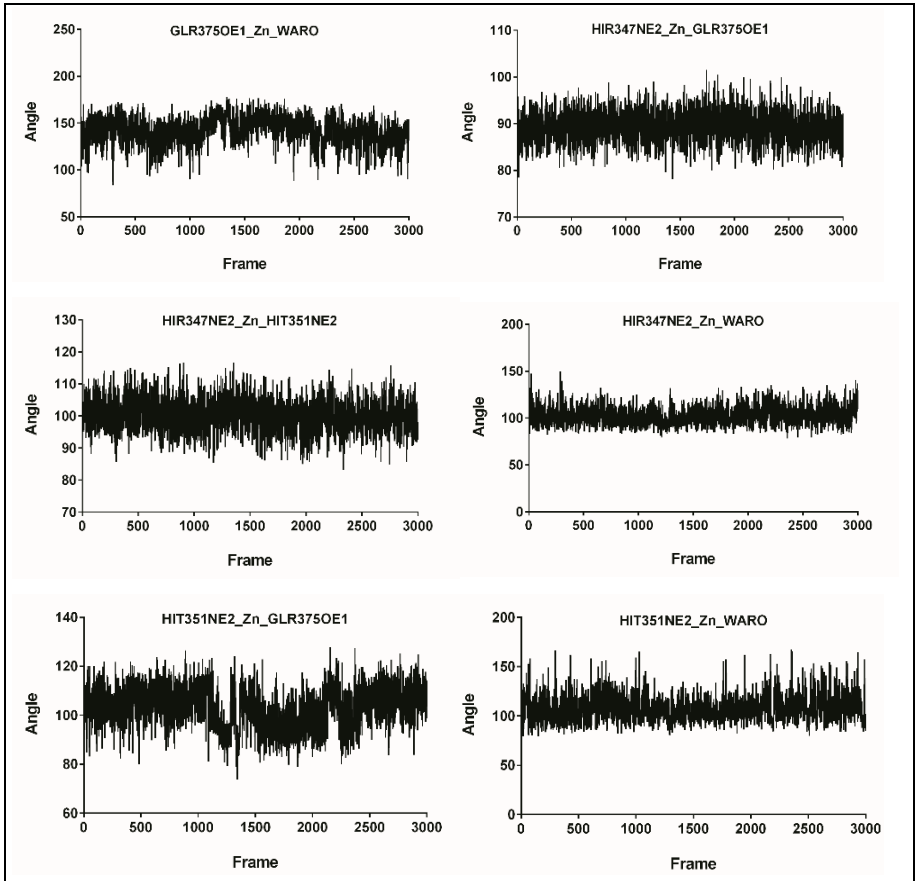


Figure A22: Variation in C-domain Zn-coordinating bond angles during 30ns simulation.

A17.5 Stability of N-domain Cl^- in the second pocket

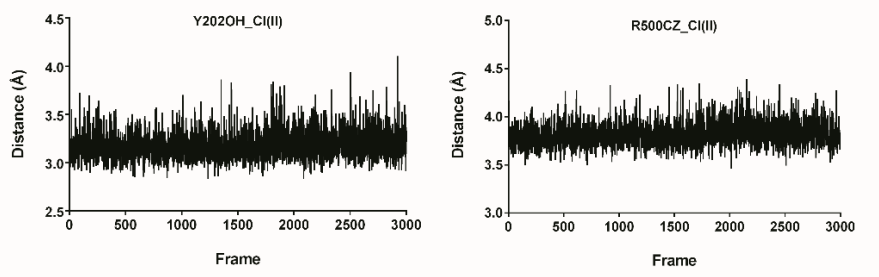


Figure A23: Variation in chloride ion binding distances in the N-domain pocket (II) during 30ns simulation.

A17.6 Stability of C-domain Cl^- in the first and second pockets

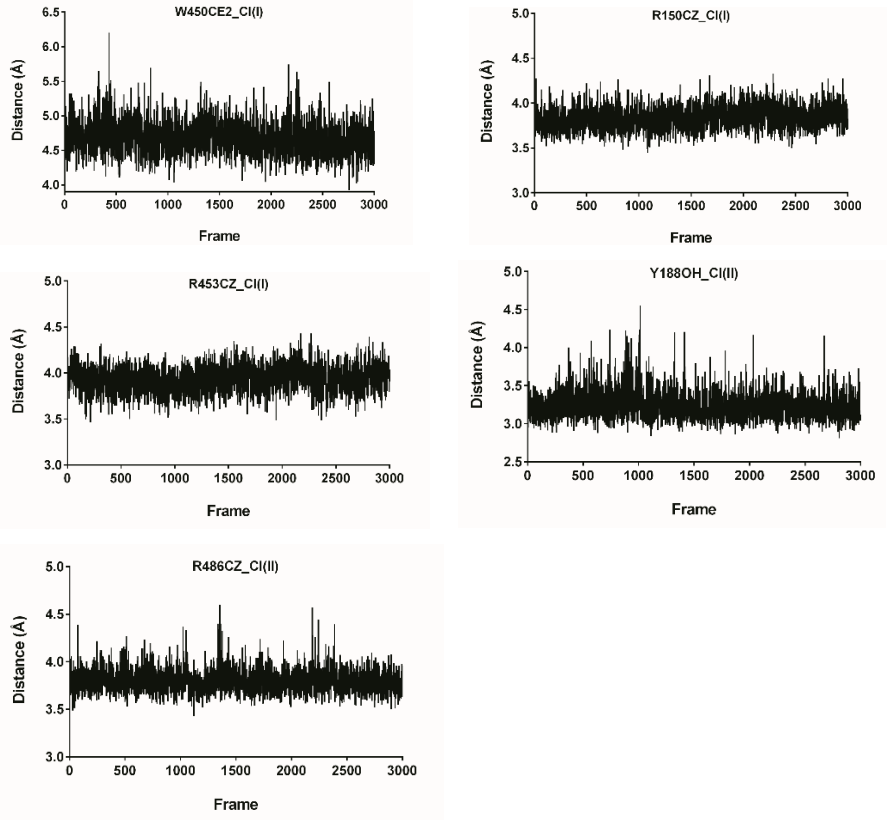


Figure A24: Variation in chloride ion binding distances in the C-domain pockets (I) and (II) during 30ns simulation.

A18 Torsion angle analysis results

A18.1 Population analysis of N-domain glycosidic torsion angles

Table A3: Population of N-domain Asn-GlcNAc linkage torsion angles during MD.
Angles represented as mean ± standard deviation with population in parentheses

	PNGs	Torsion angle				
		φ	ψ	χ_1	χ_2	ω
Asn - GlcNAc	9	243±15	169±11	180±9	240±15	NA
	25	248±11	165±7	174±7	254±10	
	45	237±15	170±10	181±7	196±13	
	82	250±17	177±10	305±10	297±18	
	117	246±14	176±8	201±7	194±14	
	131	251±16	176±13	195±17 (65%); 283±13 (35%)	68±16 (40%); 268±26 (60%)	
	289	265±14	180±11	188±9	165±30	
	416	255±17	184±10	191±10	189±20	
	480	271±12	179±9	201±9	195±17	

Table A4: Population of N-domain *α(1-6)* linkages' torsion angles during MD.
Angles represented as mean ± standard deviation with population in parentheses

	PNGs	Torsion angle				
		φ	***ψ***	***χ₁***	***χ₂***	***ω***
$Man_{\alpha 1\text{-}6}$ Man	9	70±11	108±46	NA		49±10 (99,97%); 346 (0,03%)
	25	71±10 (99%);137±23% (1%)	170±38			55±10% (93%); 281±12 (7%)
	45	75±11 (95%); 160±13 (5%)	190±18			58±11
	82	76±12	94±12			50±8
	117	68±11	183±17			53±10 (99,9%); 311±31 (0,1%)
	131	76±12	96±20			50±9 (99,97%); 359 (0,03%)
	289	72±10 (99%); 155±14 (1%)	189±15			55±11
	416	68±10 (95%); 165±10 (5%)	180±18			57±11 (15%); 180±12 (85%)
	480	76±14	125±38			52±9 (99,97%); 334 (0,03%)
$Fuc_{\alpha 1\text{-}6}$ GlcNAc	9	282±15	77±12 (23%); 166±15 (77%)	NA		49±10 (99,97%); 340 (0,03%)
	25	250±9	251±19			56±9
	45	256±28	73±13 (37%); 169±14 (63%)			42±12 (98%); 340±16 (2%)
	82	280±25	155±21			36±14 (99%); 343±19 (1%)
	117	242±39	171±15			43±11 (69%); 319±16 (31%)
	131	289±15	165±22			52±14 (99%); 345±13 (1%)
	289	281±15	165±23			44±11 (99,9%); 358±2 (0,1%)
	416	283±17	160±40			47±12 (99%); 324±16 (1%)
	480	289±10	196±13			55±11 (99,97%); 353 (0,03%)

Table A5: Population of N-domain *α(2-3)* linkages' torsion angles during MD.

Angles represented as mean ± standard deviation with population in parentheses ([a]: NeuAc on $Man_{\alpha 1\text{-}3}$ Man arm; [b]: NeuAc on $Man_{\alpha 1\text{-}6}$ Man arm)

	PNGs	Torsion angle				
		φ	***ψ***	***χ₁***	***χ₂***	***ω***
$Gal_{\alpha 2\text{-}3}$ NeuAc[a]	9	52±13 (61%); 313±13 (39%)	211±23	NA		NA
	25	56±13 (96%); 313±12 (4%)	228±27			
	45	57±14 (85%); 313±13 (15%)	221±26			
	82	57±13 (84%); 312±13 (17%)	221±26			
	117	57±14 (76%); 313±12 (24%)	221±27			
	131	56±14 (88%); 313±15 (12%)	220±27			
	289	58±14 (91%); 314±12 (9%)	221±25			
	416	54±16 (37%); 314±11 (63%)	217±18			
	480	53±14 (70%); 313±12 (30%)	214±24			
$Gal_{\alpha 2\text{-}3}$ NeuAc[b]	9	61±14 (81%); 313±12 (19%)	230±26	NA		NA
	25	59±14 (96%); 311±14 (4%)	221±27			
	45	64±12 (80%); 312±11 (20%)	235±23			
	82	54±14 (95%); 318±13 (5%)	224±25			
	117	69±8	252±8			
	131	58±14 (86%); 313±11 (14%)	222±29			
	289	55±13 (80%); 314±13 (20%)	213±21			
	416	53±12 (99,97%); 348 (0,03%)	215±28			
	480	54±15 (63%); 312±11 (37%)	212±23			

A18.2 Population analysis of C-domain glycosidic torsion angles

Table A6: Population of C-domain Asn-GlcNAc linkage torsion angles during MD.
Angles represented as mean ± standard deviation with population in parentheses

PNGs		Torsion angle				
		φ	ψ	χ_1	χ_2	ω
Asn - GlcNAc	36	227±12	172±10	177±7	199±11	NA
	54	251±16	177±10	189±9	199±19	
	73	266±14	187±9	182±7	130±16	
	119	258±17	173±10	203±11	255±31	
	301	255±17	177±10	191±11 (57%); 282±12 (43%)	215±26	
	550	247±17	173±9	207±9	89±21	

Table A7: Population of C-domain *α(1-6)* linkages' torsion angles during MD.
Angles represented as mean ± standard deviation with population in parentheses.

PNGs		Torsion angle				
		φ	***ψ***	***χ₁***	***χ₂***	***ω***
$Man_{\alpha 1\text{-}6}$Man	36	68±15	182±14	NA		50±10 (99,6%); 295±12 (0,4%)
	54	74±13	181±14			57±11 (99,9%); 297±35 (0,1%)
	73	78±13	211±20			59±13
	119	76±13	107±22			53±9
	301	67±11	67±12 (85%); 172±17 (15%)			43±11 (99%); 309±14 (1%)
	550	79±25	184±19			55±11 (99,9%); 315±48 (0,1%)
$Fuc_{\alpha 1\text{-}6}$ GlcNAc	36	287±9	196±10	NA		49±9
	54	289±10	158±22			49±11 (99,8%); 353±6 (0,2%)
	73	291±12	213±24			53±12 (99%); 342±17 (1%)
	119	277±22	171±16			42±13 (99,6%); 334±24 (0,4%)
	301	286±16	171±33			48±12 (99%); 312±27 (1%)
	550	279±19	138±44			50±10

Table A8: Population of C-domain *α(2-3)* linkages' torsion angles during MD.

Angles represented as mean ± standard deviation with population in parentheses. ([a]: NeuAc on $Man_{\alpha 1\text{-}3}$ Man arm; [b]: NeuAc on $Man_{\alpha 1\text{-}6}$ Man arm)

	PNGs	Torsion angle				
		φ	***ψ***	***χ₁***	***χ₂***	***ω***
$Gal_{\alpha 2\text{-}3}$ NeuAc[a]	36	62±15 (87%); 313±11 (13%)	228±27	NA		NA
	54	58±13 (91%); 315±11 (9%)	225±26			
	73	58±18 (89%); 312±12 (11%)	219±30			
	119	56±13 (87%); 316±12 (13%)	220±26			
	301	59±14 (80%); 313±11 (20%)	224±27			
	550	60±14 (80%); 314±12 (20%)	226±25			
$Gal_{\alpha 2\text{-}3}$ NeuAc[b]	36	55±12 (99,9%); 344±15 (0,1%)	213±22	NA		NA
	54	58±12	220±25			
	73	61±12	231±22			
	119	74±14	247±20			
	301	66±15 (91%); 316±11 (9%)	233±24			
	550	60±14 (85%); 310±14 (15%)	225±27			

A19 Hydrogen bond analyses between glycans and the N- or C-domain

Table A9: Glycan to N-domain hydrogen bonds.
Hydrogen bonds present in more than 5% of the simulation are shown with the frequencies indicated in parentheses and acceptor residue names italicized.

Donor glycan	Donor sugar - *acceptor residue* (% frames)
n2	0SA O4 - *Q30 OE1* (13.9) 3LB O4 - *Q545 OE1* (7.5) UYB C1 - *R380 HH22* (5.7) 0SA O4 - *Q371 OE1* (5.5)
n3	0SA O9 - *E49 OE1* (27.0) 0fA O2 - *N50 OD1* (24.8) 0SA O8 - *E49 OE2* (18.7) 0SA O9 - *E49 OE2* (18.4) 0SA O8 - *E49 OE1* (8.7)
n4	0SA O8 - *N203 OD1* (5.9)
n5	0SA O7 - *T482 OG1* (51.2)
n7	0SA O4 - *G316 O* (33.5) 4YB O6 - *E315 OE2* (14.6) 4YB O6 - *E315 OE1* (13.9)
n8	4YB N2 - *E522 OE2* (22.0) VMB O4 - *E522 OE1* (20.1) 0SA O8 - *Q575 OE1* (16.2) 3LB O4 - *E518 OE1* (14.4) 0SA O4 - *V560 O* (14.0) 3LB O6 - *E518 OE2* (11.7) 3LB O4 - *E518 OE2* (11.3) 3LB O6 - *E518 OE1* (10.8) 3LB O2 - *Q575 OE1* (10.5) 0SA O9 - *E518 OE2* (9.6) VMB O4 - *E522 OE2* (8.6) VMB C1 - *E522 HB3* (5.8)

0fA = Fuc, 0SA = NeuAc, 2MA = Man, 3LB = Gal, 4YB = GlcNAc, UYB =1st core GlcNAc, VMB = 1st Man linked to core GlcNAc

Table A10: N-domain to glycan hydrogen bonds.
Hydrogen bonds present in more than 5% of the simulation are shown with the frequencies indicated in parentheses and acceptor residue names italicized.

Acceptor glycan	*Donor residue* – acceptor sugar (% frames)
n2	*R340 NH2* - 4YB O2N (70.4) *R340 NH1* - 4YB O2N (36.6) *Q22 NE2* - UYB O2N (20.9) *R541 NH2* - 0SA O1A (7.1) *R541 NH2* - 0SA O1B (5.6)
n3	*R52 NH1* - 0SA O1A (14.3) *R52 NH2* - 0SA O1A (12.4) *N50 ND2* - UYB O5 (10.6) *R53 NH1* - 0fA O4 (8.5) *R52 NH1* - 0SA O1B (6.2) *R52 NH1* - 0SA O9 (5.7) *R52 NH2* - 0SA O1B (5.1)
n4	*R89 NH2* - 0fA O4 (17.4) *N203 ND2* - 0SA O1B (13.4) *N203 ND2* -0SAO1A (7.5) *R89 NH1* - 0fA O4 (7.4) *R89 NH2* - 0SA O1A (5.3) *R89 NH1* - 0SA O1A (5.2)
n5	*N480 ND2* - 0SA O4 (54.7) *H483 NE2* - 3LB O4 (41.4) *N480 ND2* - 0SA H4O (27.7) *H483 NE2* - 3LB H4O (11.6)
n7	*R344 NH1* - 0SA O4 (34.2) *R344 NH1* - 0SA O1B (11.5) *R344 NH2* - 0SA O1A (10.9) *R344 NH1* - 0SA O1A (9.0) *R344 NH2* - 0SA O1B (7.4)
n8	*Q575 NE2* - 0SA O7 (25.0) *Q579 NE2* - 0SA O7 (13.1) *Q575 NE2* - 0SA O7 (11.8) *Q568 NE2* - 4YB O4 (7.2)

0fA = Fuc, 0SA = NeuAc, 2MA = Man, 3LB = Gal, 4YB = GlcNAc, UYB =1st core GlcNAc, VMB = 1st Man linked to core GlcNAc

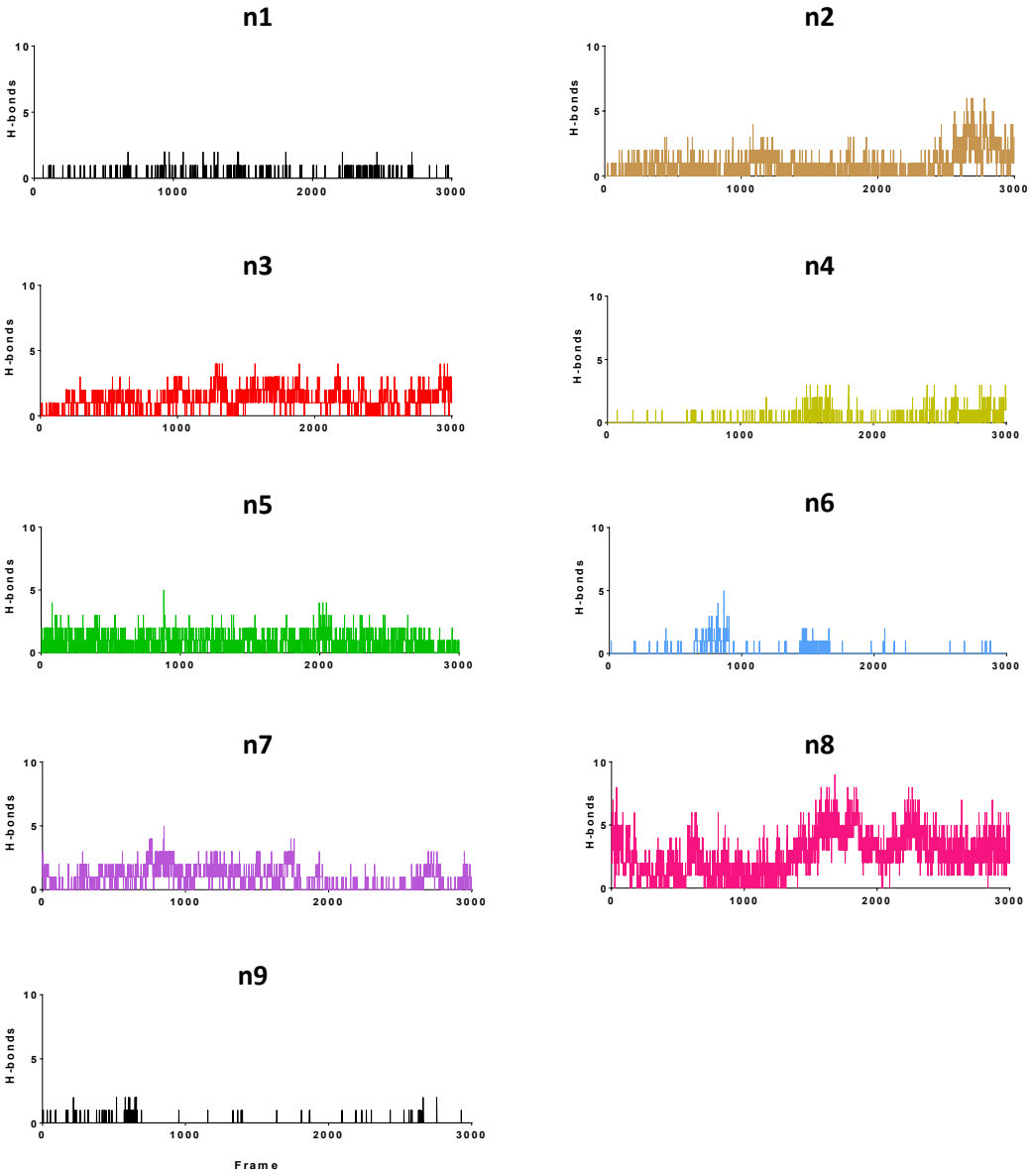


Figure A25: Hydrogen bond frequency between each glycan and the N-domain.

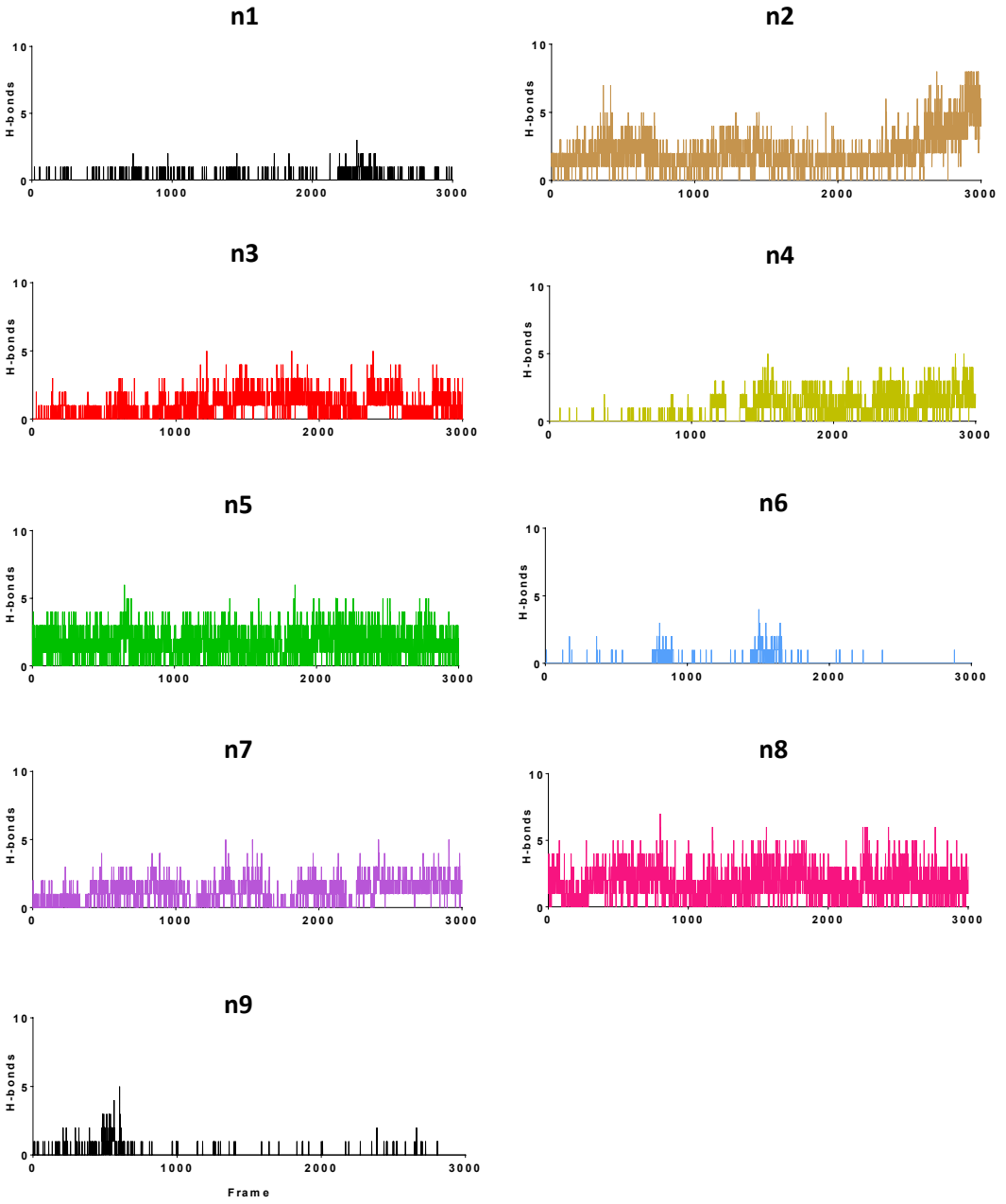


Figure A26: Hydrogen bond frequency between the N-domain and each glycan.

Table A11: Glycan to C-domain hydrogen bonds.
Hydrogen bonds present in more than 5% of the simulation are shown with the frequencies indicated in parentheses and acceptor residue names italicized.

Donor glycan	Donor sugar - acceptor residue (% frames)
c1	0SA O7 - *E25 OE2* (35.4) 0SA O7 - *E25 OE1* (27.7) 0SA O8 - *E25 OE2* (8.7) 0SA O8 - *E25 OE1* (5.9) 0SA O4 - *Q47 OE1* (5.2)
c2	0SA O8 - *E14 OE2* (14.2) 0SA O8 - *E14 OE1* (14.2) 0SA O9 - *E14 OE1* (9.7) 0SA O4 - *E62 OE1* (6.6)
c3	0SA O4 - *N325 O* (63.8) 0fA O4 - *E13 OE2* (35.6) 0fA O4 - *E13 OE1* (29.3) 0SA O9 - *D360 OD1* (9.2)
c4	0SA O4 - *G120 O* (27.0)
c5	0fA O4 - *E28 OE2* (6.7)
c6	UYB O3 - *S192 OG* (37.7) 4YB O6 - *S192 OG* (35.2) UYB N2 - *D196 OD1* (29.9) UYB O3 - *S192 HG* (18.5) UYB O3 - *D196 OD1* (12.8) UYB N2 - *D196 OD2* (12.1) UYB O3 - *D196 OD2* (11.1)

0fA = Fuc, 0SA = NeuAc, 2MA = Man, 3LB = Gal, 4YB = GlcNAc, UYB =1st core GlcNAc, VMB = 1st Man linked to core GlcNAc

Table A12: C-domain to glycan hydrogen bonds.
Hydrogen bonds present in more than 5% of the simulation are shown with the frequencies indicated in parentheses and acceptor residue names italicized.

Acceptor glycan	Donor residue – acceptor sugar (% frames)
c1	*K48 NZ* - 0SA O5N (15.9) *Q47 NE2* - 0SA O4 (14.1) *Q51 NE2* - 0SA O4 (13.7) *K48 NZ* - 0SA O5N (13.0) *K48 NZ* - 0SA O5N (12.9) *Q51 NE2* - 0SA O4 (9.8)
c2	*Q51 NE2* - 0fA O4 (10.0) *K58 NZ* - 0SA O7 (7.8) *Q62 NE2* - 0SA O5N (7.4) *K58 NZ* - 0SA O7 (6.4) *K58 NZ* - 0SA O7 (6.0)
c4	*C122 N* - 0SA O5N (21.8)
c5	*W31 NE1* - 0fA O4 (9.0)
c6	*M551 N* - UYB O2N (13.9) *S192 N* - 4YB H6O (13.3) *S192 N* - 4YB O6 (12.9)

0fA = Fuc, 0SA = NeuAc, 2MA = Man, 3LB = Gal, 4YB = GlcNAc, UYB =1st core GlcNAc, VMB = 1st Man linked to core GlcNAc

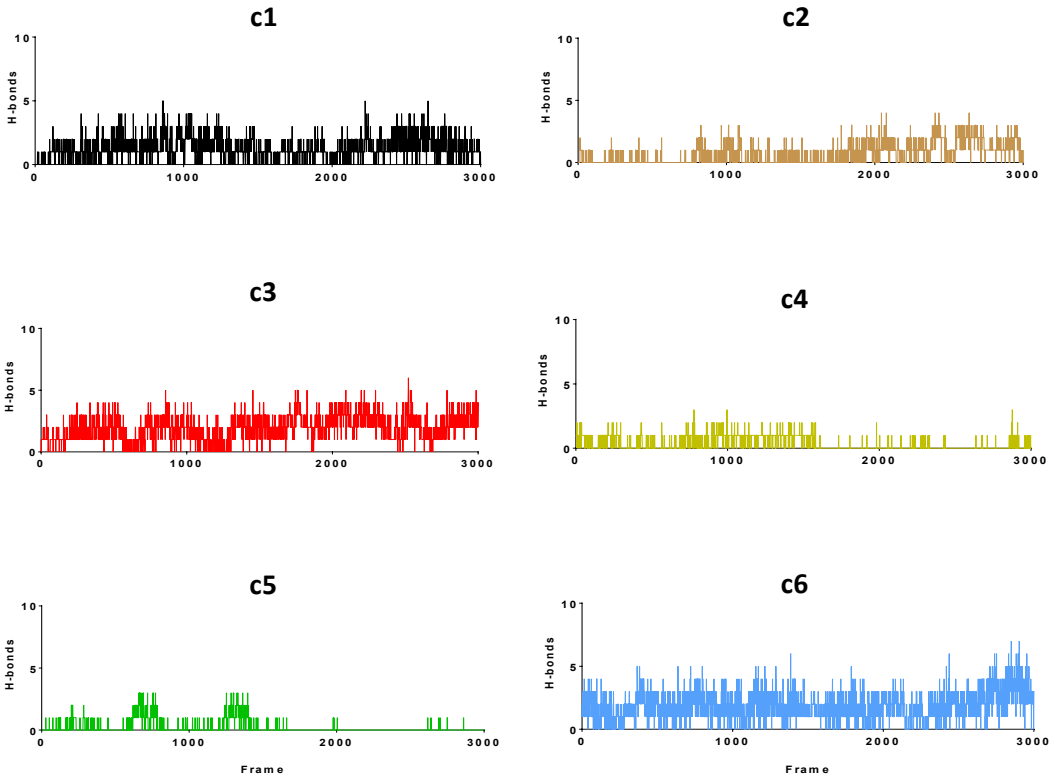


Figure A27: Hydrogen bond frequency between each glycan and the C-domain.

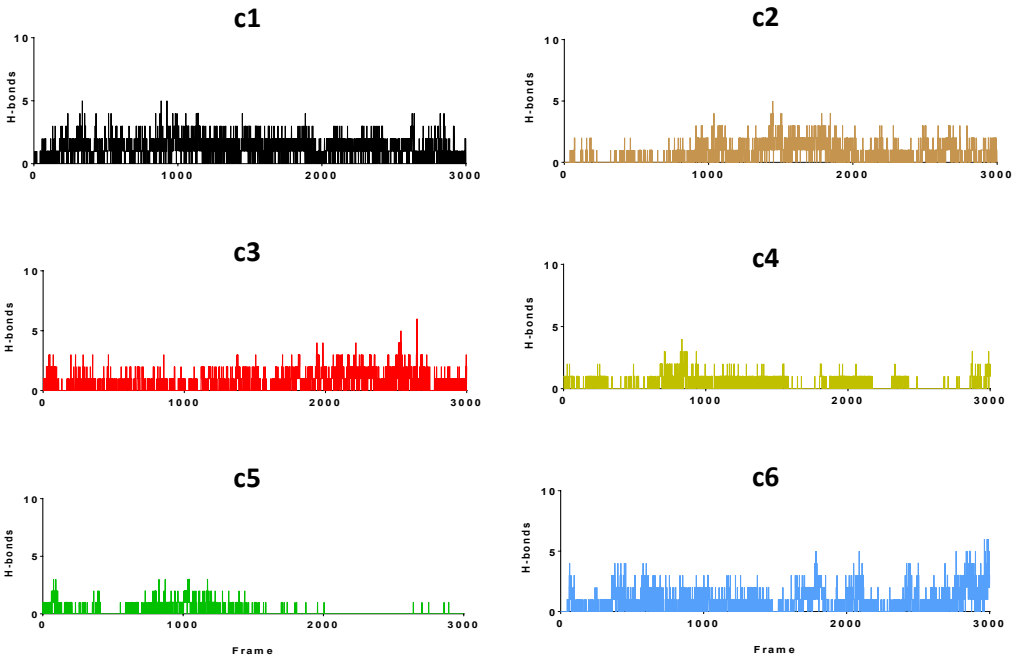


Figure A28: Hydrogen bond frequency between the C-domain and each glycan.

A20 AQUA-DUCT statistics

Table A13: Statistical analysis of the N-domain water flow.
Entry and exit paths are described by rows and columns, respectively, with the number of water molecules given for each. Water molecules that do not exit but remain in the scope region for the duration of simulation are given in red.

Incoming path \ *Outgoing path*	Prime	Non-prime	Flap 1	Flap 2	Scope	Outliers	Sum
Prime	11 (28.21%)	1 (2.56%)	3 (7.69%)	2 (5.13%)	**2** (5.13%)		19 (48.72%)
Non-prime	3 (7.69%)			1 (2.56%)	**5** (12.82%)	1 (2.56%)	10 (25.64%)
Flap 1	1 (2.56%)		1 (2.56%)		**2** (5.13%)		4 (10.28%)
Flap 2	1 (2.56%)	1 (2.56%)					2 (5.13%)
Scope	1 (2.56%)						1 (2.56%)
Outliers	1 (2.56%)				**1** (2.56%)	1 (2.56%)	3 (7.69%)
Sum	18 (46.15%)	2 (5.13%)	4 (10.26%)	3 (7.69%)	**10** (25.64%)	2 (5.13%)	39

Table A14: Statistical analysis of the C-domain water flow.
Entry and exit paths are described by rows and columns, respectively, with the number of water molecules given for each. Water molecules that do not exit but remain in the scope region for the duration of simulation are given in red.

Incoming path \ *Outgoing path*	Prime	Non-prime	Flap 1	Flap 2	Scope	Outliers	Sum
Prime	31 (16.23%)	18 (9.42%)		3 (1.57%)	**28** (14.66%)	3 (1.57%)	83 (43.46%)
Non-prime	20 (10.47%)	9 (4.71%)	1 (0.52%)	1 (0.52%)	**16** (8.38%)	2 (1.05%)	49 (25.65%)
Flap 1	1 (0.52%)						1 (0.52%)
Flap 2	4 (2.09%)	1 (0.52%)				1 (0.52%)	6 (3.14%)
Scope	27 (14.14%)	13 (6.81%)	1 (0.52%)	1 (0.52%)	**1** (0.52%)	5 (2.62%)	48 (25.13%)
Outliers	1 (0.52%)	2 (1.05%)				1 (0.52%)	4 (2.09%)
Sum	84 (44.0%)	43 (22.51%)	2 (1.05%)	5 (2.62%)	**45** (23.56%)	12 (6.28%)	191

References

ABU HASAN, Z., WILLIAMS, H., ISMAIL, N. M., OTHMAN, H., COZIER, G. E., ACHARYA, K. R. & ISAAC, R. E. 2017. The toxicity of angiotensin converting enzyme inhibitors to larvae of the disease vectors Aedes aegypti and Anopheles gambiae. *Sci Rep,* 7**,** 45409.

ACHARYA, K. R., STURROCK, E. D., RIORDAN, J. F. & EHLERS, M. R. 2003. Ace revisited: a new target for structure-based drug design. *Nature reviews.Drug discovery,* 2**,** 891-902.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

AGARWAL, R. P. & PERRIN, D. D. 1977. Copper(II) and zinc(II) complexes of glycylglycyl-L-histidine and derivatives. *Journal of the Chemical Society, Dalton Transactions***,** 53-57.

AGBALE, C. M., CARDOSO, M. H., GALYUON, I. K. & FRANCO, O. L. 2016. Designing metallodrugs with nuclease and protease activity. *Metallomics,* 8**,** 1159-1169.

AHLUWALIA, A. & PERRETTI, M. 1999. B1 receptors as a new inflammatory target. Could this B the 1? *Trends Pharmacol Sci,* 20**,** 100-4.

ANTHONY, C. S. 2011. *The Importance of N-linked Glycosylation on the N-domain of Angiotensin-I Converting Enzyme.* Doctor of Philosophy (PhD), University of Cape Town.

ANTHONY, C. S., CORRADI, H. R., SCHWAGER, S. L., REDELINGHUYS, P., GEORGIADIS, D., DIVE, V., ACHARYA, K. R. & STURROCK, E. D. 2010. The N domain of human angiotensin-I-converting enzyme: the role of N-glycosylation and the crystal structure in complex with an N domain-specific phosphinic inhibitor, RXP407. *The Journal of biological chemistry,* 285**,** 35685-35693.

ARAUJO, M. C., MELO, R. L., CESARI, M. H., JULIANO, M. A., JULIANO, L. & CARMONA, A. K. 2000. Peptidase specificity characterization of C- and N-terminal catalytic sites of angiotensin I-converting enzyme. *Biochemistry,* 39**,** 8519-25.

AUSUBEL, F. M., BRENT, R., KINGSTON, R. E., MOORE, D. D., SEIDMAN, J. G., SMITH, J. A. & STRUHL, K. 1989. *Short Protocols in Molecular Biology,* New York, Greene Publishing Associates.

BECKMAN, J. S. & KOPPENOL, W. H. 1996. Nitric oxide, superoxide, and peroxynitrite: the good, the bad, and ugly. *Am J Physiol,* 271**,** C1424-37.

BERSANETTI, P. A., ANDRADE, M. C., CASARINI, D. E., JULIANO, M. A., NCHINDA, A. T., STURROCK, E. D., JULIANO, L. & CARMONA, A. K. 2004. Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides for defining substrate specificity of the angiotensin I-converting enzyme and development of selective C-domain substrates. *Biochemistry,* 43**,** 15729-36.

BICKET, D. P. 2002. Using ACE inhibitors appropriately. *American Family Physician,* 66**,** 461-468.

BINEVSKI, P. V., SIZOVA, E. A., POZDNEV, V. F. & KOST, O. A. 2003. Evidence for the negative cooperativity of the two active sites within bovine somatic angiotensin-converting enzyme. *FEBS Lett,* 550**,** 84-8.

BINKOWSKI, T. A., NAGHIBZADEH, S. & LIANG, J. 2003. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Research,* 31**,** 3352-3355.

BIOVIA, D. S. 2013. Discovery Studio Modeling Environment. San Diego: Dassault Systèmes.

BITTERMAN, P. B. & HENKE, C. A. 1991. Fibroproliferative disorders. *Chest,* 99**,** 81s-84s.

BOLTE, E., VERDY, M., MARC-AURELE, J., BROUILLET, J., BEAUREGARD, P. & GENEST, J. 1958. Studies on new diuretic compounds: spirolactone and chlorothiazide. *Canadian Medical Association journal,* 79**,** 881-888.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* 72**,** 248-254.

BRADFORD, S. & COWAN, J. A. 2012. Catalytic metallodrugs targeting HCV IRES RNA. *Chemical Communications,* 48**,** 3118-3120.

BRÁS, N. F., FERNANDES, P. A. & RAMOS, M. J. 2014. QM/MM Study and MD Simulations on the Hypertension Regulator Angiotensin-Converting Enzyme. *ACS Catalysis,* 4**,** 2587-2597.

BROWN, N. J., GAINER, J. V., MURPHEY, L. J. & VAUGHAN, D. E. 2000. Bradykinin stimulates tissue plasminogen activator release from human forearm vasculature through B-2 receptor-dependent, NO synthase-independent, and cyclooxygenase-independent pathway. *Circulation,* 102**,** 2190-2196.

BROWN, N. J. & VAUGHAN, D. E. 1998. Angiotensin-converting enzyme inhibitors. *Circulation,* 97**,** 1411-20.

BULL, H. G., THORNBERRY, N. A. & CORDES, E. H. 1985. Purification of angiotensin-converting enzyme from rabbit lung and human plasma by affinity chromatography. *The Journal of biological chemistry,* 260**,** 2963-2972.

BURNHAM, S., SMITH, J. A., LEE, A. J., ISAAC, R. E. & SHIRRAS, A. D. 2005. The angiotensin-converting enzyme (ACE) gene family of Anopheles gambiae. *BMC Genomics,* 6**,** 172.

CAMERMAN, N., CAMERMAN, A. & SARKAR, B. 1976. Molecular design to mimic the copper(II) transport site of human albumin. The crystal and molecular structure of copper(II) – glycylglycyl-L-histidine-N-methyl amide monoaquo complex. *Canadian Journal of Chemistry,* 54**,** 1309-1316.

CANDIANO, G., BRUSCHI, M., MUSANTE, L., SANTUCCI, L., GHIGGERI, G. M., CARNEMOLLA, B., ORECCHIA, P., ZARDI, L. & RIGHETTI, P. G. 2004. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis,* 25**,** 1327-33.

CARMONA, A. K., SCHWAGER, S. L., JULIANO, M. A., JULIANO, L. & STURROCK, E. D. 2006. A continuous fluorescence resonance energy transfer angiotensin I-converting enzyme assay. *Nat Protoc,* 1**,** 1971-6.

CERONI, A., MAASS, K., GEYER, H., GEYER, R., DELL, A. & HASLAM, S. M. 2008. GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *Journal of Proteome Research,* 7**,** 1650-1659.

CHAE, P. S., KIM, M.-S., JEUNG, C.-S., LEE, S. D., PARK, H., LEE, S. & SUH, J. 2005. Peptide-Cleaving Catalyst Selective for Peptide Deformylase. *Journal of the American Chemical Society,* 127**,** 2396-2397.

CHEI, W. S., JU, H. & SUH, J. 2011. New chelating ligands for Co(III)-based peptide-cleaving catalysts selective for pathogenic proteins of amyloidoses. *J Biol Inorg Chem,* 16**,** 511-9.

CHEN, H. L., LUNSDORF, H., HECHT, H. J. & TSAI, H. 2010. Porcine pulmonary angiotensin I-converting enzyme--biochemical characterization and spatial arrangement of the N- and C-domains by three-dimensional electron microscopic reconstruction. *Micron,* 41**,** 674-85.

CHEN, X. & CATRAVAS, J. D. 1993. Neutrophil-mediated endothelial angiotensin-converting enzyme dysfunction: role of oxygen-derived free radicals. *Am J Physiol,* 265**,** L243-9.

CHEN, X., TZANELA, M., BAUMGARTNER, M. K., MCCORMICK, J. R. & CATRAVAS, J. D. 1992. PMA-activated neutrophils decrease ectoenzyme activities in rabbit aortic endothelial cells in culture. *Am J Physiol,* 263**,** L657-63.

CHENG, Y. & PRUSOFF, W. H. 1973. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol,* 22**,** 3099-108.

CHOO, H. Y., YOON, H. R., PARK, H. S., KIM, D. H., PARK, J. & KIM, D. H. 1998. Epoxyalkanoyls as novel ACE inhibitors. *Arch Pharm Res,* 21**,** 168-73.

COMPANY, C., PIQUERAS, L., NAIM ABU NABAH, Y., ESCUDERO, P., BLANES, J. I., JOSE, P. J., MORCILLO, E. J. & SANZ, M.-J. 2011. Contributions of ACE and mast cell chymase to endogenous angiotensin II generation and leucocyte recruitment in vivo. *Cardiovascular Research,* 92**,** 48-56.

CORRADI, H. R., SCHWAGER, S. L., NCHINDA, A. T., STURROCK, E. D. & ACHARYA, K. R. 2006. Crystal structure of the N domain of human somatic angiotensin I-converting enzyme provides a structural basis for domain-specific inhibitor design. *Journal of Molecular Biology,* 357**,** 964-974.

COTTON, J., HAYASHI, M. A., CUNIASSE, P., VAZEUX, G., IANZER, D., DE CAMARGO, A. C. & DIVE, V. 2002. Selective inhibition of the C-domain of angiotensin I converting enzyme by bradykinin potentiating peptides. *Biochemistry,* 41**,** 6065-71.

COWAN, J. A. 2008. Catalytic metallodrugs. *Pure and Applied Chemistry,* 80**,** 1799-1810.

COZIER, G. E., SCHWAGER, S. L., SHARMA, R. K., CHIBALE, K., STURROCK, E. D. & ACHARYA, K. R. 2018. Crystal structures of sampatrilat and sampatrilat-Asp in complex with human ACE - a molecular basis for domain selectivity. *Febs j,* 285**,** 1477-1490.

CUENOUD, B., TARASOW, T. M. & SCHEPARTZ, A. 1992. A new strategy for directed protein cleavage. *Tetrahedron Letters,* 33**,** 895-898.

DANAEI, G., DING, E. L., MOZAFFARIAN, D., TAYLOR, B., REHM, J., MURRAY, C. J. & EZZATI, M. 2009. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med,* 6**,** e1000058.

DARDEN, T., YORK, D. & PEDERSEN, L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics,* 98**,** 10089-10092.

DE GASPARO, M., CATT, K. J., INAGAMI, T., WRIGHT, J. W. & UNGER, T. 2000. International union of pharmacology. XXIII. The angiotensin II receptors. *Pharmacol Rev,* 52**,** 415-72.

DEDDISH, P. A., MARCIC, B., JACKMAN, H. L., WANG, H. Z., SKIDGEL, R. A. & ERDOS, E. G. 1998. N-domain-specific substrate and C-domain inhibitors of angiotensin-converting enzyme: angiotensin-(1-7) and keto-ACE. *Hypertension,* 31**,** 912-7.

DEDDISH, P. A., WANG, J., MICHEL, B., MORRIS, P. W., DAVIDSON, N. O., SKIDGEL, R. A. & ERDOS, E. G. 1994. Naturally occurring active N-domain of human angiotensin I-converting enzyme. *Proc Natl Acad Sci U S A,* 91**,** 7807-11.

DENTI, P., SHARP, S. K., KROGER, W. L., SCHWAGER, S. L., MAHAJAN, A., NJOROGE, M., GIBHARD, L., SMIT, I., CHIBALE, K., WIESNER, L., STURROCK, E. D. & DAVIES, N. H. 2014. Pharmacokinetic evaluation of lisinopril-tryptophan, a novel C-domain ACE inhibitor. *Eur J Pharm Sci,* 56**,** 113-9.

DIVE, V., COTTON, J., YIOTAKIS, A., MICHAUD, A., VASSILIOU, S., JIRACEK, J., VAZEUX, G., CHAUVET, M. T., CUNIASSE, P. & CORVOL, P. 1999. RXP 407, a phosphinic peptide, is a potent inhibitor of angiotensin I converting enzyme able to differentiate between its two active sites. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 4330-4335.

DIVE, V., GEORGIADIS, D., MATZIARI, M., MAKARITIS, A., BEAU, F., CUNIASSE, P. & YIOTAKIS, A. 2004. Phosphinic peptides as zinc metalloproteinase inhibitors. *Cellular and molecular life sciences : CMLS,* 61**,** 2010-2019.

DIXON, M. 1953. The determination of enzyme inhibitor constants. *The Biochemical journal,* 55**,** 170-171.

DOUGLAS, R. G. 2011. *Significance of Active Site Residues in the N-Domain Selectivity of Angiotensin-Converting Enzyme.* Doctor of Philosophy, University of Cape Town.

DOUGLAS, R. G., SHARMA, R. K., MASUYER, G., LUBBE, L., ZAMORA, I., ACHARYA, K. R., CHIBALE, K. & STURROCK, E. D. 2014. Fragment-based design for the development of N-domain-selective angiotensin-1-converting enzyme inhibitors. *Clin Sci (Lond),* 126**,** 305-13.

DU BOIS, R. M. 2011. Idiopathic pulmonary fibrosis: present understanding and future options. *European respiratory review : an official journal of the European Respiratory Society,* 20**,** 132-133.

EHLERS, M. R., FOX, E. A., STRYDOM, D. J. & RIORDAN, J. F. 1989. Molecular cloning of human testicular angiotensin-converting enzyme: the testis isozyme is identical to the C-terminal half of endothelial angiotensin-converting enzyme. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 7741-7745.

EHLERS, M. R., GORDON, K., SCHWAGER, S. L. & STURROCK, E. D. 2012. Shedding the load of hypertension: the proteolytic processing of angiotensin-converting enzyme. *S Afr Med J,* 102**,** 461-4.

EHLERS, M. R. W., CHEN, Y. P. & RIORDAN, J. F. 1992. The unique N-terminal sequence of testis angiotensin-converting enzyme is heavily O-glycosylated and unessential for activity or stability. *Biochemical and biophysical research communications,* 183**,** 199-205.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

ERDOS, E. G. 1990. Angiotensin-I Converting Enzyme and the Changes in Our Concepts through the Years - Dahl,Lewis,K. Memorial Lecture. *Hypertension,* 16**,** 363-370.

FELLER, S. E., ZHANG, Y., PASTOR, R. W. & BROOKS, B. R. 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics,* 103**,** 4613-4621.

FIDAI, I., HOCHAROEN, L., BRADFORD, S., WACHNOWSKY, C. & COWAN, J. A. 2014. Inactivation of sortase A mediated by metal ATCUN complexes. *J Biol Inorg Chem,* 19**,** 1327-39.

FIGUEROA, C. D., MATUS, C. E., PAVICIC, F., SARMIENTO, J., HIDALGO, M. A., BURGOS, R. A., GONZALEZ, C. B., BHOOLA, K. D. & EHRENFELD, P. 2015. Kinin B1 receptor regulates interactions between neutrophils and endothelial cells by modulating the levels of Mac-1, LFA-1 and intercellular adhesion molecule-1. *Innate Immun,* 21**,** 289-304.

FOGARI, R. & ZOPPI, A. 2006. Antihypertensive drugs and fibrinolytic function. *Am J Hypertens,* 19**,** 1293-9.

FRIEDLAND, J. & SILVERSTEIN, E. 1976. A sensitive fluorimetric assay for serum angiotensin-converting enzyme. *Am J Clin Pathol,* 66**,** 416-24.

FUCHS, S. 2004. Role of the N-terminal catalytic domain of angiotensin-converting enzyme investigated by targeted inactivation in mice. *Journal of Biological Chemistry,* 279**,** 15946.

FUCHS, S., XIAO, H. D., HUBERT, C., MICHAUD, A., CAMPBELL, D. J., ADAMS, J. W., CAPECCHI, M. R., CORVOL, P. & BERNSTEIN, K. E. 2008. Angiotensin-converting enzyme C-terminal catalytic domain is the main site of angiotensin I cleavage in vivo. *Hypertension,* 51**,** 267-274.

GALLAGHER, J., ZELENKO, O., WALTS, A. D. & SIGMAN, D. S. 1998. Protease activity of 1,10-phenanthroline-copper(I). Targeted scission of the catalytic site of carbonic anhydrase. *Biochemistry,* 37**,** 2096-104.

GALM, U., HAGER, M. H., VAN LANEN, S. G., JU, J., THORSON, J. S. & SHEN, B. 2005. Antitumor antibiotics: bleomycin, enediynes, and mitomycin. *Chem Rev,* 105**,** 739-58.

GASYMOV, O. K. & GLASGOW, B. J. 2007. ANS fluorescence: potential to augment the identification of the external binding sites of proteins. *Biochim Biophys Acta,* 1774**,** 403-11.

GEORGIADIS, D., BEAU, F., CZARNY, B., COTTON, J., YIOTAKIS, A. & DIVE, V. 2003. Roles of the two active sites of somatic angiotensin-converting enzyme in the cleavage of angiotensin I and bradykinin: insights from selective inhibitors. *Circulation research,* 93**,** 148-154.

GEORGIADIS, D., CUNIASSE, P., COTTON, J., YIOTAKIS, A. & DIVE, V. 2004. Structural determinants of RXPA380, a potent and highly selective inhibitor of the angiotensin-converting enzyme C-domain. *Biochemistry,* 43**,** 8048-8054.

GHOSH, A. K. & VAUGHAN, D. E. 2012. PAI-1 in tissue fibrosis. *J Cell Physiol,* 227**,** 493-507.

GOH, J. B. & NG, S. K. 2017. Impact of host cell line choice on glycan profile. *Critical Reviews in Biotechnology***,** 1-17.

GOKHALE, N. H. & COWAN, J. A. 2005. Inactivation of human angiotensin converting enzyme by copper peptide complexes containing ATCUN motifs. *Chem Commun (Camb)***,** 5916-8.

GOKHALE, N. H. & COWAN, J. A. 2006. Metallopeptide-promoted inactivation of angiotensin-converting enzyme and endothelin-converting enzyme 1: Toward dual-action therapeutics. *J Biol Inorg Chem,* 11**,** 937-47.

GONZALEZ, P., BOSSAK, K., STEFANIAK, E., HUREAU, C., RAIBAUT, L., BAL, W. & FALLER, P. 2018. N-Terminal Cu-Binding Motifs (Xxx-Zzz-His, Xxx-His) and Their Derivatives: Chemistry, Biology and Medicinal Applications. *Chemistry – A European Journal,* 0.

GORDON, K. 2011. *Protein-protein interactions of human somatic angiotensin-converting enzyme.* Doctor of Philosophy (PhD), University of Cape Town.

GORDON, K., REDELINGHUYS, P., SCHWAGER, S. L., EHLERS, M. R., PAPAGEORGIOU, A. C., NATESH, R., ACHARYA, K. R. & STURROCK, E. D. 2003. Deglycosylation, processing and crystallization of human testis angiotensin-converting enzyme. *Biochem J,* 371**,** 437-42.

GRADY, E. F., SECHI, L. A., GRIFFIN, C. A., SCHAMBELAN, M. & KALINYAK, J. E. 1991. Expression of AT2 receptors in the developing rat fetus. *The Journal of clinical investigation,* 88**,** 921-933.

GRANT, B. J., RODRIGUES, A. P., ELSAWY, K. M., MCCAMMON, J. A. & CAVES, L. S. 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics,* 22**,** 2695-6.

HAATAJA, L., GURLO, T., HUANG, C. J. & BUTLER, P. C. 2008. Islet Amyloid in Type 2 Diabetes, and the Toxic Oligomer Hypothesis. *Endocrine Reviews,* 29**,** 303-316.

HAGAMAN, J. R., MOYER, J. S., BACHMAN, E. S., SIBONY, M., MAGYAR, P. L., WELCH, J. E., SMITHIES, O., KREGE, J. H. & O'BRIEN, D. A. 1998. Angiotensin-converting enzyme and male fertility. *Proc Natl Acad Sci U S A,* 95**,** 2552-7.

HAIT, S. K. & MOULIK, S. P. 2001. Determination of critical micelle concentration (CMC) of nonionic surfactants by donor-acceptor interaction with Iodine and correlation of CMC with hydrophile-lipophile balance and other parameters of the surfactants. *Journal of Surfactants and Detergents,* 4**,** 303-309.

HARFORD, C. & SARKAR, B. 1997. Amino terminal Cu(II)- and Ni(II)-binding (ATCUN) motif of proteins and peptides: Metal binding, DNA cleavage, and other properties. *Accounts of Chemical Research,* 30**,** 123-130.

HARRISON, C. & ACHARYA, K. R. 2014. ACE for all - a molecular perspective. *J Cell Commun Signal,* 8**,** 195-210.

HELENIUS, A., MCCASLIN, D. R., FRIES, E. & TANFORD, C. 1979. Properties of detergents. *Methods Enzymol,* 56**,** 734-49.

HINES, C. S., RAY, K., SCHMIDT, J. J., XIONG, F., FEENSTRA, R. W., PRAS-RAVES, M., DE MOES, J. P., LANGE, J. H., MELIKISHVILI, M., FRIED, M. G., MORTENSON, P., CHARLTON, M., PATEL, Y., COURTNEY, S. M., KRUSE, C. G. & RODGERS, D. W. 2014. Allosteric inhibition of the neuropeptidase neurolysin. *J Biol Chem,* 289**,** 35605-19.

HOANG, M. V. & TURNER, A. J. 1997. Novel activity of endothelin-converting enzyme: hydrolysis of bradykinin. *Biochemical Journal,* 327**,** 23-26.

HOCHAROEN, L., JOYNER, J. C. & COWAN, J. A. 2013. N- versus C-Domain Selectivity of Catalytic Inactivation of Human Angiotensin Converting Enzyme by Lisinopril-Coupled Transition Metal Chelates. *Journal of Medicinal Chemistry,* 56**,** 9826-9836.

HOOVER, W. G. 1985. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A,* 31**,** 1695-1697.

HOYER, D., CHO, H. & SCHULTZ, P. G. 1990. New strategy for selective protein cleavage. *Journal of the American Chemical Society,* 112**,** 3249-3250.

HUANG, W., RAVIKUMAR, K. M., CHANCE, M. R. & YANG, S. 2015. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis. *Biophys J,* 108**,** 107-15.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics,* 14**,** 33-38.

HUTCHINSON, J., FOGARTY, A., HUBBARD, R. & MCKEEVER, T. 2015. Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J,* 46**,** 795-806.

ISCHIROPOULOS, H. & AL-MEHDI, A. B. 1995. Peroxynitrite-mediated oxidative protein modifications. *FEBS Lett,* 364**,** 279-82.

JASPARD, E., WEI, L. & ALHENC-GELAS, F. 1993. Differences in the properties and enzymatic specificities of the two active sites of angiotensin I-converting enzyme (kininase II). Studies with bradykinin and other natural peptides. *The Journal of biological chemistry,* 268**,** 9496-9503.

JIN, Y. & COWAN, J. A. 2005. DNA cleavage by copper-ATCUN complexes. Factors influencing cleavage mechanism and linearization of dsDNA. *J Am Chem Soc,* 127**,** 8408-15.

JIN, Y. & COWAN, J. A. 2006. Targeted Cleavage of HIV Rev Response Element RNA by Metallopeptide Complexes. *Journal of the American Chemical Society,* 128**,** 410-411.

JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W. & KLEIN, M. L. 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics,* 79**,** 926-935.

JOYNER, J. C. 2012. *Synthesis and Evaluation of Catalytic Metallodrugs and Analysis of RNA Cleavage by Mass Spectrometry.* Doctor of Philosophy (PhD0, Ohio State University.

JOYNER, J. C. & COWAN, J. A. 2011. Targeted cleavage of HIV RRE RNA by Rev-coupled transition metal chelates. *J Am Chem Soc,* 133**,** 9912-22.

JOYNER, J. C., HOCHAROEN, L. & COWAN, J. A. 2012. Targeted catalytic inactivation of angiotensin converting enzyme by lisinopril-coupled transition-metal chelates. *J Am Chem Soc,* 134**,** 3396-410.

JUNOT, C., GONZALES, M. F., EZAN, E., COTTON, J., VAZEUX, G., MICHAUD, A., AZIZI, M., VASSILIOU, S., YIOTAKIS, A., CORVOL, P. & DIVE, V. 2001. RXP 407, a selective inhibitor of the N-domain of angiotensin I-converting enzyme, blocks in vivo the degradation of hemoregulatory peptide acetyl-Ser-Asp-Lys-Pro with no effect on angiotensin I hydrolysis. *The Journal of pharmacology and experimental therapeutics,* 297**,** 606-611.

KATSANOS, G. S., ANOGEIANAKI, A., ORSO, C., TETE, S., SALINI, V., ANTINOLFI, P. L. & SABATINO, G. 2008. Impact of substance P on cellular immunity. *J Biol Regul Homeost Agents,* 22**,** 93-8.

KESSLER, S. P., ROWE, T. M., GOMOS, J. B., KESSLER, P. M. & SEN, G. C. 2000. Physiological non-equivalence of the two isoforms of angiotensin-converting enzyme. *The Journal of biological chemistry,* 275**,** 26259-26264.

KIM, H. J., HA, S., LEE, H. Y. & LEE, K. J. 2015. ROSics: chemistry and proteomics of cysteine modifications in redox biology. *Mass Spectrom Rev,* 34**,** 184-208.

KIM, M. G., KIM, M. S., LEE, S. D. & SUH, J. 2006. Peptide-cleaving catalyst selective for melanin-concentrating hormone: Oxidative decarboxylation of N-terminal aspartate catalyzed by Co(III)cyclen. *J Biol Inorg Chem,* 11**,** 867-75.

KIM, M. S., JEON, J. W. & SUH, J. 2005. Angiotensin-cleaving catalysts: conversion of N-terminal aspartate to pyruvate through oxidative decarboxylation catalyzed by Co(III)cyclen. *J Biol Inorg Chem,* 10**,** 364-72.

KIRSCHNER, K. N., YONGYE, A. B., TSCHAMPEL, S. M., GONZALEZ-OUTEIRINO, J., DANIELS, C. R., FOLEY, B. L. & WOODS, R. J. 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem,* 29**,** 622-55.

KOHARA, H., TAJIMA, S., YAMAMOTO, M. & TABATA, Y. 2010. Angiogenesis induced by controlled release of neuropeptide substance P. *Biomaterials,* 31**,** 8617-25.

KOHLSTEDT, K., BRANDES, R. P., MULLER-ESTERL, W., BUSSE, R. & FLEMING, I. 2004. Angiotensin-converting enzyme is involved in outside-in signaling in endothelial cells. *Circ Res,* 94**,** 60-7.

KOHLSTEDT, K., GERSHOME, C., FRIEDRICH, M., MULLER-ESTERL, W., ALHENC-GELAS, F., BUSSE, R. & FLEMING, I. 2006. Angiotensin-converting enzyme (ACE) dimerization is the initial step in the ACE inhibitor-induced ACE signaling cascade in endothelial cells. *Mol Pharmacol,* 69**,** 1725-32.

KOZLIKOVA, B., SEBESTOVA, E., SUSTR, V., BREZOVSKY, J., STRNAD, O., DANIEL, L., BEDNAR, D., PAVELKA, A., MANAK, M., BEZDEKA, M., BENES, P., KOTRY, M., GORA, A., DAMBORSKY, J. & SOCHOR, J. 2014. CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics,* 30**,** 2684-5.

KROGER, W. L., DOUGLAS, R. G., O'NEILL, H. G., DIVE, V. & STURROCK, E. D. 2009. Investigating the domain specificity of phosphinic inhibitors RXPA380 and RXP407 in angiotensin-converting enzyme. *Biochemistry,* 48**,** 8405-8412.

KRYUKOVA, O. V., TIKHOMIROVA, V. E., GOLUKHOVA, E. Z., EVDOKIMOV, V. V., KALANTAROV, G. F., TRAKHT, I. N., SCHWARTZ, D. E., DULL, R. O., GUSAKOV, A. V., UPOROV, I. V., KOST, O. A. & DANILOV, S. M. 2015. Tissue Specificity of Human Angiotensin I-Converting Enzyme. *PLOS ONE,* 10**,** e0143455.

KUMAR, S., DIETRICH, N. & KORNFELD, K. 2016. Angiotensin Converting Enzyme (ACE) Inhibitor Extends Caenorhabditis elegans Life Span. *PLoS Genetics,* 12**,** e1005866.

LABINJOH, C., NEWBY, D. E., PELLEGRINI, M. P., JOHNSTON, N. R., BOON, N. A. & WEBB, D. J. 2001. Potentiation of bradykinin-induced tissue plasminogen activator release by angiotensin-converting enzyme inhibition. *J Am Coll Cardiol,* 38**,** 1402-8.

LAEMMLI, U. K. 1970. SDS-page Laemmli method. *Nature,* 227**,** 680.

LAU, S.-J., KRUCK, T. P. A. & SARKAR, B. 1974. A Peptide Molecule Mimicking the Copper(II) Transport Site of Human Serum Albumin: A COMPARATIVE STUDY BETWEEN THE SYNTHETIC SITE AND ALBUMIN. *Journal of Biological Chemistry,* 249**,** 5878-5884.

LAU, S.-J. & SARKAR, B. 1971. Ternary Coordination Complex between Human Serum Albumin, Copper (II), and l-Histidine. *Journal of Biological Chemistry,* 246**,** 5938-5943.

LEE, H. S., QI, Y. & IM, W. 2015. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Scientific Reports,* 5**,** 8926.

LEE, J., HO YOO, S., JEONG, K., YEON LEE, T., YOUNG AHN, J. & SUH, J. 2008. *Cleavage agents for alpha-synuclein*.

LI, X., GRANT, O. C., ITO, K., WALLACE, A., WANG, S., ZHAO, P., WELLS, L., LU, S., WOODS, R. J. & SHARP, J. S. 2017. Structural Analysis of the Glycosylated Intact HIV-1 gp120–b12 Antibody Complex Using Hydroxyl Radical Protein Footprinting. *Biochemistry,* 56**,** 957-970.

LIAO, T.-D., YANG, X.-P., D'AMBROSIO, M., ZHANG, Y., RHALEB, N.-E. & CARRETERO, O. A. 2010. N-Acetyl-Seryl-Aspartyl-Lysyl-Proline Attenuates Renal Injury and Dysfunction

in Hypertensive Rats With Reduced Renal Mass: Council for High Blood Pressure Research. *Hypertension,* 55**,** 459-467.

LIN, L. & HU, K. 2014. Tissue plasminogen activator and inflammation: from phenotype to signaling mechanisms. *American Journal of Clinical and Experimental Immunology,* 3**,** 30-36.

LIN, L. & HU, K. 2017. Tissue-type plasminogen activator modulates macrophage M2 to M1 phenotypic change through annexin A2-mediated NF-κB pathway. *Oncotarget,* 8**,** 88094-88103.

LIN, L., JIN, Y. & HU, K. 2015. Tissue-type Plasminogen Activator (tPA) Promotes M1 Macrophage Survival through p90 Ribosomal S6 Kinase (RSK) and p38 Mitogen-activated Protein Kinase (MAPK) Pathway. *The Journal of Biological Chemistry,* 290**,** 7910-7917.

LINKE, D. 2009. Detergents: an overview. *Methods Enzymol,* 463**,** 603-17.

LINLEY, J. E., OOI, L., PETTINGER, L., KIRTON, H., BOYLE, J. P., PEERS, C. & GAMPER, N. 2012. Reactive oxygen species are second messengers of neurokinin signaling in peripheral sensory neurons. *Proceedings of the National Academy of Sciences,* 109**,** 9246.

LIU, Y., KATI, W., CHEN, C. M., TRIPATHI, R., MOLLA, A. & KOHLBRENNER, W. 1999. Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. *Anal Biochem,* 267**,** 331-5.

MADHAVI SASTRY, G., ADZHIGIREY, M., DAY, T., ANNABHIMOJU, R. & SHERMAN, W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design,* 27**,** 221-234.

MAGDZIARZ, T., MITUSINSKA, K., GOLDOWSKA, S., PLUCIENNIK, A., STOLARCZYK, M., LUGOWSKA, M. & GORA, A. 2017. AQUA-DUCT: a ligands tracking tool. *Bioinformatics,* 33**,** 2045-2046.

MAIER, J. A., MARTINEZ, C., KASAVAJHALA, K., WICKSTROM, L., HAUSER, K. E. & SIMMERLING, C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation,* 11**,** 3696-3713.

MCDONALD, M. R., FREDERICKS, F. C. & MARGERUM, D. W. 1997. Characterization of Copper(III)−Tetrapeptide Complexes with Histidine as the Third Residue. *Inorganic Chemistry,* 36**,** 3119-3124.

MICHEL, B., GRIMA, M., NIRINA, L. B., INGERT, C., COQUARD, C., BARTHELMEBS, M. & IMBS, J. L. 2001. Inhibitory effect of reactive oxygen species on angiotensin I-converting enzyme (kininase II). *Clin Exp Pharmacol Physiol,* 28**,** 212-8.

MILLS, K. T., BUNDY, J. D., KELLY, T. N., REED, J. E., KEARNEY, P. M., REYNOLDS, K., CHEN, J. & HE, J. 2016. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. *Circulation,* 134**,** 441-50.

MITRA, N., SINHA, S., RAMYA, T. N. C. & SUROLIA, A. 2006. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends in Biochemical Sciences,* 31**,** 156-163.

MOMANY, F. A. & RONE, R. 1992. Validation of the general purpose QUANTA ®3.2/CHARMm® force field. *Journal of Computational Chemistry,* 13**,** 888-900.

NAKAGAWA, K., KITAZUME, S., OKA, R., MARUYAMA, K., SAIDO, T. C., SATO, Y., ENDO, T. & HASHIMOTO, Y. 2006. Sialylation enhances the secretion of neurotoxic amyloid-beta peptides. *J Neurochem,* 96**,** 924-33.

NATESH, R., SCHWAGER, S. L., STURROCK, E. D. & ACHARYA, K. R. 2003. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature,* 421**,** 551-554.

NCHINDA, A. T., CHIBALE, K., REDELINGHUYS, P. & STURROCK, E. D. 2006a. Synthesis and molecular modeling of a lisinopril-tryptophan analogue inhibitor of angiotensin I-converting enzyme. *Bioorg Med Chem Lett,* 16**,** 4616-9.

NCHINDA, A. T., CHIBALE, K., REDELINGHUYS, P. & STURROCK, E. D. 2006b. Synthesis of novel keto-ACE analogues as domain-selective angiotensin I-converting enzyme inhibitors. *Bioorg Med Chem Lett,* 16**,** 4612-5.

NIESEN, F. H., BERGLUND, H. & VEDADI, M. 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc,* 2**,** 2212-21.

NKOE, K. M. 2014. *The Role of N-linked Glycosylation on the Structure and Function of Somatic Angiotensin-Converting Enzyme.* Masters of Science (MSc), University of Cape Town.

NOSÉ, S. 1984. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics,* 81**,** 511-519.

O'NEILL, H. G., REDELINGHUYS, P., SCHWAGER, S. L. & STURROCK, E. D. 2008. The role of glycosylation and domain interactions in the thermal stability of human angiotensin-converting enzyme. *Biol Chem,* 389**,** 1153-61.

OBA, R., IGARASHI, A., KAMATA, M., NAGATA, K., TAKANO, S. & NAKAGAWA, H. 2005. The N-terminal active centre of human angiotensin-converting enzyme degrades Alzheimer amyloid beta-peptide. *The European journal of neuroscience,* 21**,** 733-740.

ONDETTI, M. A., RUBIN, B. & CUSHMAN, D. W. 1977. Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents. *Science (New York, N.Y.),* 196**,** 441-444.

OTZEN, D. 2011. Protein–surfactant interactions: A tale of many states. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1814**,** 562-591.

PAPAKYRIAKOU, A., SPYROULIAS, G. A., STURROCK, E. D., MANESSI-ZOUPA, E. & CORDOPATIS, P. 2007. Simulated interactions between angiotensin-converting enzyme and substrate gonadotropin-releasing hormone: novel insights into domain selectivity. *Biochemistry,* 46**,** 8753-65.

PARK CHOO, H. Y., PEAK, K. H., PARK, J., KIM, D. H. & CHUNG, H. S. 2000. Design and synthesis of alpha,beta-unsaturated carbonyl compounds as potential ACE inhibitors. *Eur J Med Chem,* 35**,** 643-8.

PARK, H. I., LEE, S., ULLAH, A., CAO, Q. & SANG, Q. X. 2010. Effects of detergents on catalytic activity of human endometase/matrilysin 2, a putative cancer biomarker. *Anal Biochem,* 396**,** 262-8.

PATCHETT, A. A., HARRIS, E., TRISTRAM, E. W., WYVRATT, M. J., WU, M. T., TAUB, D., PETERSON, E. R., IKELER, T. J., TEN BROEKE, J., PAYNE, L. G., ONDEYKA, D. L., THORSETT, E. D., GREENLEE, W. J., LOHR, N. S., HOFFSOMMER, R. D., JOSHUA, H., RUYLE, W. V., ROTHROCK, J. W., ASTER, S. D., MAYCOCK, A. L., ROBINSON, F. M., HIRSCHMANN, R., SWEET, C. S., ULM, E. H., GROSS, D. M., VASSIL, T. C. & STONE, C. A. 1980. A new class of angiotensin-converting enzyme inhibitors. *Nature,* 288**,** 280-3.

PENG, H., CARRETERO, O. A., BRIGSTOCK, D. R., OJA-TEBBE, N. & RHALEB, N.-E. 2003. Ac-SDKP Reverses Cardiac Fibrosis in Rats With Renovascular Hypertension. *Hypertension,* 42**,** 1164-1170.

PENG, H., CARRETERO, O. A., VULJAJ, N., LIAO, T. D., MOTIVALA, A., PETERSON, E. L. & RHALEB, N. E. 2005. Angiotensin-converting enzyme inhibitors: a new mechanism of action. *Circulation,* 112**,** 2436-2445.

PETRESCU, A. J., MILAC, A. L., PETRESCU, S. M., DWEK, R. A. & WORMALD, M. R. 2004. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology,* 14**,** 103-14.

PETRESCU, A. J., PETRESCU, S. M., DWEK, R. A. & WORMALD, M. R. 1999. A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. *Glycobiology,* 9**,** 343-52.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem,* 25**,** 1605-12.

PRICE, D. J. & BROOKS, C. L., 3RD 2004. A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys,* 121**,** 10096-103.

RASOUL, S., CARRETERO, O. A., PENG, H., CAVASIN, M. A., ZHUO, J., SANCHEZ-MENDOZA, A., BRIGSTOCK, D. R. & RHALEB, N. E. 2004. Antifibrotic effect of Ac-SDKP and angiotensin-converting enzyme inhibition in hypertension. *Journal of hypertension,* 22**,** 593-603.

REDELINGHUYS, P. 2006. *Structure-Function Relationship of Angiotensin-Converting Enzyme: Glycosylation and Domain-Selectivity.* Doctor of Philosophy (PhD), University of Cape Town.

REDELINGHUYS, P., NCHINDA, A. T. & STURROCK, E. D. 2005. Development of domain-selective angiotensin I-converting enzyme inhibitors. *Annals of the New York Academy of Sciences,* 1056**,** 160-175.

REILLY, C. F., TEWKSBURY, D. A., SCHECHTER, N. M. & TRAVIS, J. 1982. Rapid conversion of angiotensin I to angiotensin II by neutrophil and mast cell proteinases. *J Biol Chem,* 257**,** 8619-22.

RIVIERE, G., FELLOUS, A., FRANCO, A., BERNAY, B. & FAVREL, P. 2011. A crucial role in fertility for the oyster angiotensin-converting enzyme orthologue CgACE. *PLoS One,* 6**,** e27833.

ROUSSEAU, A., MICHAUD, A., CHAUVET, M. T., LENFANT, M. & CORVOL, P. 1995. The hemoregulatory peptide N-acetyl-Ser-Asp-Lys-Pro is a natural and specific substrate of the N-terminal active site of human angiotensin-converting enzyme. *The Journal of biological chemistry,* 270**,** 3656-3661.

SALI, A. & BLUNDELL, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol,* 234**,** 779-815.

SANTOS, E. L., DE PICOLI SOUZA, K., DA SILVA, E. D., BATISTA, E. C., MARTINS, P. J., D'ALMEIDA, V. & PESQUERO, J. B. 2009. Long term treatment with ACE inhibitor enalapril decreases body weight gain and increases life span in rats. *Biochem Pharmacol,* 78**,** 951-8.

SASTRY, G. M., ADZHIGIREY, M., DAY, T., ANNABHIMOJU, R. & SHERMAN, W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des,* 27**,** 221-34.

SCHARFSTEIN, J., RAMOS, P. I. P. & BARRAL-NETTO, M. 2017. G Protein-Coupled Kinin Receptors and Immunity Against Pathogens. *Adv Immunol,* 136**,** 29-84.

SCHEPARTZ, A. & CUENOUD, B. 1990. Site-specific cleavage of the protein calmodulin using a trifluoperazine-based affinity reagent. *Journal of the American Chemical Society,* 112**,** 3247-3249.

SCHIFFMANN, E., CORCORAN, B. A. & WAHL, S. M. 1975. N-formylmethionyl peptides as chemoattractants for leucocytes. *Proceedings of the National Academy of Sciences of the United States of America,* 72**,** 1059-1062.

SCHRODINGER, LLC 2017. The PyMOL Molecular Graphics System, Version 2.0.6.

SCHWAGER, S. L., CARMONA, A. K. & STURROCK, E. D. 2006. A high-throughput fluorimetric assay for angiotensin I-converting enzyme. *Nature Protocols,* 1**,** 1961-1964.

SHARMA, J. N., AL-OMRAN, A. & PARVATHY, S. S. 2007. Role of nitric oxide in inflammatory diseases. *Inflammopharmacology,* 15**,** 252-9.

SHARMA, R. K., DOUGLAS, R. G., LOUW, S., CHIBALE, K. & STURROCK, E. D. 2012. New ketomethylene inhibitor analogues: synthesis and assessment of structural determinants for N-domain selective inhibition of angiotensin-converting enzyme. *Biol Chem,* 393**,** 485-93.

SHARMA, U., RHALEB, N. E., POKHAREL, S., HARDING, P., RASOUL, S., PENG, H. & CARRETERO, O. A. 2008. Novel anti-inflammatory mechanisms of N-Acetyl-Ser-Asp-Lys-Pro in hypertension-induced target organ damage. *Am J Physiol Heart Circ Physiol,* 294**,** H1226-32.

SHENKIN, P. S., YARMUSH, D. L., FINE, R. M., WANG, H. J. & LEVINTHAL, C. 1987. Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. *Biopolymers,* 26**,** 2053-85.

SHRIVASTAVA, A. 2007. Activation of macrophages with N-formyl-methionyl-leucyl-phenylalanine: involvement of protein kinase C and tyrosine kinase. *Indian J Exp Biol,* 45**,** 755-63.

SIGEL, H. & MARTIN, R. B. 1982. Coordinating properties of the amide bond. Stability and structure of metal ion complexes of peptides and related ligands. *Chemical Reviews,* 82**,** 385-426.

SIMOES E SILVA, A. C., SILVEIRA, K. D., FERREIRA, A. J. & TEIXEIRA, M. M. 2013. ACE2, angiotensin-(1-7) and Mas receptor axis in inflammation and fibrosis. *Br J Pharmacol,* 169**,** 477-92.

SIRAGY, H. M. 2000. The role of the AT2 receptor in hypertension. *Am J Hypertens,* 13**,** 62s-67s.

SKIDGEL, R. A. & ERDOS, E. G. 1985. Novel activity of human angiotensin I converting enzyme: release of the NH2- and COOH-terminal tripeptides from the luteinizing hormone-releasing hormone. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 1025-1029.

SKIRGELLO, O. E., BINEVSKI, P. V., POZDNEV, V. F. & KOST, O. A. 2005. Kinetic probes for inter-domain co-operation in human somatic angiotensin-converting enzyme. *Biochem J,* 391**,** 641-7.

SOUBRIER, F. 1988. Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning. *Proceedings of the National Academy of Sciences of the United States of America,* 85**,** 9386.

STURROCK, E. D., NATESH, R., VAN ROOYEN, J. M. & ACHARYA, K. R. 2004. Structure of angiotensin I-converting enzyme. *Cell Mol Life Sci,* 61**,** 2677-86.

STURROCK, E. D., YU, X. C., WU, Z., BIEMANN, K. & RIORDAN, J. F. 1996. Assignment of Free and Disulfide-Bonded Cysteine Residues in Testis Angiotensin-Converting Enzyme: Functional Implications. *Biochemistry,* 35**,** 9560-9566.

SUGIYAMA, H., KILKUSKIE, R. E., CHANG, L. H., MA, L. T., HECHT, S. M., VAN DER MAREL, G. A. & VAN BOOM, J. H. 1986. DNA strand scission by bleomycin: catalytic cleavage and strand selectivity. *Journal of the American Chemical Society,* 108**,** 3852-3854.

SUH, J., CHEI, W. S., LEE, T. Y., KIM, M. G., YOO, S. H., JEONG, K. & AHN, J. Y. 2008. Cleavage agents for soluble oligomers of human islet amyloid polypeptide. *J Biol Inorg Chem,* 13**,** 693-701.

SUH, J., YOO, S. H., KIM, M. G., JEONG, K., AHN, J. Y., KIM, M. S., CHAE, P. S., LEE, T. Y., LEE, J., LEE, J., JANG, Y. A. & KO, E. H. 2007. Cleavage agents for soluble oligomers of amyloid beta peptides. *Angew Chem Int Ed Engl,* 46**,** 7064-7.

THEOHARIDES, T. C., ALYSANDRATOS, K.-D., ANGELIDOU, A., DELIVANIS, D.-A., SISMANOPOULOS, N., ZHANG, B., ASADI, S., VASIADI, M., WENG, Z., MINIATI, A. & KALOGEROMITROS, D. 2012. Mast cells and inflammation. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1822**,** 21-33.

TOMINEY, A. F., DOCHERTY, P. H., ROSAIR, G. M., QUENARDELLE, R. & KRAFT, A. 2006. Unusually Weak Binding Interactions in Tetrazole−Amidine Complexes. *Organic Letters,* 8**,** 1279-1282.

TOWLER, P., STAKER, B., PRASAD, S. G., MENON, S., TANG, J., PARSONS, T., RYAN, D., FISHER, M., WILLIAMS, D., DALES, N. A., PATANE, M. A. & PANTOLIANO, M. W. 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J Biol Chem,* 279**,** 17996-8007.

UMEZAWA, H., MAEDA, K., TAKEUCHI, T. & OKAMI, Y. 1966. New antibiotics, bleomycin A and B. *J Antibiot (Tokyo),* 19**,** 200-9.

VARKI, A., CUMMINGS, R. D., AEBI, M., PACKER, N. H., SEEBERGER, P. H., ESKO, J. D., STANLEY, P., HART, G., DARVILL, A., KINOSHITA, T., PRESTEGARD, J. J., SCHNAAR, R. L., FREEZE, H. H., MARTH, J. D., BERTOZZI, C. R., ETZLER, M. E., FRANK, M., VLIEGENTHART, J. F., LUTTEKE, T., PEREZ, S., BOLTON, E., RUDD, P., PAULSON, J., KANEHISA, M., TOUKACH, P., AOKI-KINOSHITA, K. F., DELL, A., NARIMATSU, H., YORK, W., TANIGUCHI, N. & KORNFELD, S. 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology,* 25**,** 1323-4.

VORONOV, S., ZUEVA, N., ORLOV, V., ARUTYUNYAN, A. & KOST, O. 2002a. Temperature-induced selective death of the C-domain within angiotensin-converting enzyme molecule. *FEBS Letters,* 522**,** 77-82.

VORONOV, S. V., BINEVSKI, P. V., ZUEVA, N. A., PALYULIN, V. A., BASKIN, I. I., ORLOVA, M. A. & KOST, O. A. 2003. Structural and Functional Peculiarities of Homologous Domains of Angiotensin-Converting Enzyme. *Russian Journal of Bioorganic Chemistry,* 29**,** 426-433.

VORONOV, S. V., SKIRGELLO, O. E., TROSHINA, N. N., ORLOVA, M. A. & KOST, O. A. 2002b. A Hydrophobic Site on the Surface of the Angiotensin-Converting Enzyme Molecule. *Biochemistry (Moscow),* 67**,** 553-557.

WATERMEYER, J. M., KROGER, W. L., O'NEILL, H. G., SEWELL, B. T. & STURROCK, E. D. 2008. Probing the basis of domain-dependent inhibition using novel ketone inhibitors of Angiotensin-converting enzyme. *Biochemistry,* 47**,** 5942-50.

WATERMEYER, J. M., SEWELL, B. T., SCHWAGER, S. L., NATESH, R., CORRADI, H. R., ACHARYA, K. R. & STURROCK, E. D. 2006. Structure of testis ACE glycosylation mutants and evidence for conserved domain movement. *Biochemistry,* 45**,** 12654-63.

WEI, L. 1991. The two homologous domains of human angiotensin I-converting enzyme are both catalytically active. *Journal of Biological Chemistry,* 266**,** 9002.

WOOD, N. T., FADDA, E., DAVIS, R., GRANT, O. C., MARTIN, J. C., WOODS, R. J. & TRAVERS, S. A. 2013. The Influence of N-Linked Glycans on the Molecular Dynamics of the HIV-1 gp120 V3 Loop. *PLOS ONE,* 8**,** e80301.

WOODMAN, Z. L. 2003. *Characterisation of the Ectodomain Shedding of Angiotensin-Converting Enzyme.* Doctor of Philosophy (PhD), University of Cape Town.

WOODMAN, Z. L., SCHWAGER, S. L., REDELINGHUYS, P., CHUBB, A. J., VAN DER MERWE, E. L., EHLERS, M. R. & STURROCK, E. D. 2006. Homologous substitution of ACE C-domain regions with N-domain sequences: effect on processing, shedding, and catalytic properties. *Biol Chem,* 387**,** 1043-51.

WYNN, T. A. 2008. Cellular and molecular mechanisms of fibrosis. *The Journal of pathology,* 214**,** 199-210.

XU, H., YANG, F., SUN, Y., YUAN, Y., CHENG, H., WEI, Z., LI, S., CHENG, T., BRANN, D. & WANG, R. 2012. A new antifibrotic target of Ac-SDKP: inhibition of myofibroblast differentiation in rat lung with silicosis. *PloS one,* 7**,** e40301.

YANG, F., YANG, X.-P., LIU, Y.-H., XU, J., CINGOLANI, O., RHALEB, N.-E. & CARRETERO, O. A. 2004. Ac-SDKP Reverses Inflammation and Fibrosis in Rats With Heart Failure After Myocardial Infarction. *Hypertension,* 43**,** 229-236.

YANG, M., HUANG, J., SIMON, R., WANG, L. X. & MACKERELL, A. D., JR. 2017a. Conformational Heterogeneity of the HIV Envelope Glycan Shield. *Sci Rep,* 7**,** 4435.

YANG, W., RILEY, B. T., LEI, X., POREBSKI, B. T., KASS, I., BUCKLE, A. M. & MCGOWAN, S. 2017b. Generation of AMBER force field parameters for zinc centres of M1 and M17 family aminopeptidases. *J Biomol Struct Dyn***,** 1-10.

YATES, C. J., MASUYER, G., SCHWAGER, S. L., AKIF, M., STURROCK, E. D. & ACHARYA, K. R. 2014. Molecular and thermodynamic mechanisms of the chloride-dependent human angiotensin-I-converting enzyme (ACE). *J Biol Chem,* 289**,** 1798-814.

YU, X. C., STURROCK, E. D., WU, Z., BIEMANN, K., EHLERS, M. R. & RIORDAN, J. F. 1997. Identification of N-linked glycosylation sites in human testis angiotensin-converting enzyme and expression of an active deglycosylated form. *J Biol Chem,* 272**,** 3511-9.

YU, Z. & COWAN, J. A. 2017. Catalytic Metallodrugs: Substrate-Selective Metal Catalysts as Therapeutics. *Chemistry,* 23**,** 14113-14127.

YU, Z., HAN, M. & COWAN, J. A. 2015. Toward the design of a catalytic metallodrug: selective cleavage of G-quadruplex telomeric DNA by an anticancer copper-acridine-ATCUN complex. *Angew Chem Int Ed Engl,* 54**,** 1901-5.

ZOU, K., MAEDA, T., WATANABE, A., LIU, J., LIU, S., OBA, R., SATOH, Y., KOMANO, H. & MICHIKAWA, M. 2009. Abeta42-to-Abeta40- and angiotensin-converting activities in different domains of angiotensin-converting enzyme. *The Journal of biological chemistry,* 284**,** 31914-31920.